

个人简历

姓名: 苑军辉 出生日期: 1988 年 1 月
电子邮件 : Daleyjh@foxmail.com QQ: 905027814
手机: 17090120394 微信: daleyjh/17090120394
个人介绍站点: <http://resume.popdata.cn>
现公司: 北京当当网信息技术有限公司 (<https://www.dangdang.com/>)
现职位: 高级大数据开发工程师 职级: T3. 2
现居地: 北京市朝阳区管庄新村



教育经历

2007/9—2011/07: 青海大学(本科 211), 计算机科学与技术专业

英语水平: CET-6

专业技能

- 精通 Hadoop、HBase、Hive、Clickhouse 原理, 具备数据仓库分层理念;
- 擅长 Hadoop、Hbase、Hive、Clickhouse 优化和运维, 具备编写常用 Shell、Python 脚本能力;
- 熟悉 Spark、Flink、Presto 计算框架、ElasticSearch; 熟悉 Grafana+prometheus 监控组件; Hue 工具;
- 熟悉大数据周边数据库: Greenplum、Postgresql、MySQL、Redis、MongoDB 数据库;
- 具备一定的小团队管理经验;
- 了解 Web 开发组件: Rails、PHP; 前端 JQuery、Bootstrap、Echarts 等;
- 了解常用服务器组件: Nginx、Tomcat、Openresty;
- 了解数据湖组件: Delta、Hudi、Iceberg。

工作经验

2018/12—至今: 北京当当网信息技术有限公司

所在部门: 技术部-数据智能部

职位: 高级大数据开发工程师 T3. 2

● 职 责:

1. 负责共同设计新版数据仓库, 规范数据仓库分层。
2. 负责最新技术的调研: 如: Clickhouse 新版、腾讯云 Clickhouse 仓库、华为云 Clickhouse 仓库、数据湖 Delta、Hudi、Iceberg。
3. 负责数仓数据集市层和业务报表。
 - 1) 负责数据集市 DWS 层订单数据建模, 维度扩充;
 - 2) 负责用户画像、年度报告、年中报告等用户数据分析;
 - 3) 负责各部门、供应商、管理层等报表开发。
4. 负责搭建和维护当当网大数据平台 V3(Hadoop3、Hive3、HBase2 + Clickhouse + DataX/Waterdrop) 集群 (集群现有规模 100+, 日增量 500T+)。
 - 1) 负责大数据平台 V3 集群各组件节点上下线, 节点硬件问题维护;
 - 2) 负责 HDFS 参数优化、YARN 队列管理、YARN 参数优化、Hive3 On Tez 参数优化、HBase2 参数优化, Clickhouse 参数优化;

5. 负责维护当当自研调度平台 diss (基于 java), 调度资源控制, 作业机维护。
6. 负责维护当当网大数据平台 V2 (Hadoop2、Hive2) (集群规模最高 500+, 日增量 500T+)。
负责预下线的 V2 平台节点下线和 HDFS 文件清理, HDFS 稳定性保障。
7. 负责维护搜索推荐 Hadoop V3 (Hadoop3、Hive3、HBase2) 集群 (集群规模 200+)
负责保障搜索推荐 Hadoop 集群稳定性, 相关组件参数调整。

● **成 果:**

1. **移动悟空平台**

移动悟空平台是手机端查看当当全站实时浏览和销售情况的平台, 包含经营分析、用户分析、留存分析、渠道分析、商品分析、供应商分析。底层为报表+明细数据+预聚合数据, 主要运用 Clickhouse + Hive + Flink 实现。

2. **北极星平台**

北极星平台基于 davinci 开源框架二次开发, 除了自身报表配置功能外, 增加了实时大盘、可视化大屏、用户标签等功能。其中实时大盘为小时粒度数据, 为日常活动使用, 可视化大屏为公司大屏显示, 近实时显示, 每年书香节、618、双 11 等大促活动用。底层为报表+明细数据, 主要运用 Clickhouse + Hive + Flink 实现。

3. **UMT 平台 (Users Marketing Tool) 用户营销平台**

UMT 平台主要用户为市场营销, 主要功能为用户各维度的销售数据分布、用户群提取和用户画像。对于已知的用户特征, 根据特征组合与或非关系, 提交为 Hive 任务, 提取出相应特征的用户群, 用做市场营销。底层直接调用 hadoop 集群, 主要运用 Java + Hive + HBase 实现。

4. **大数据平台 V2 到大数据平台 V3 的优化**

- 1) 从无管理方式改为用 Ambari 管理集群, 采用 HDP3 版本, Hadoop、YARN、Hive、HBase 均启用 HA, 保障集群随时都是高可用状态。
- 2) 优化 Hadoop 参数, 在集群大规模插入分区数据和大规模删除数据的时候保证 NameNode 稳定。
- 3) YARN 从公平调度改为容量调度, 队列资源重新规划, 主要分为三类: 实时队列, 离线独立资源队列, 离线共享资源队列, 通过队列分离, 控制独占资源和共享资源。
- 4) Hadoop3 YARN 集群异构, 利用 YARN 异构方式, 扩充 Hadoop3 集群的 YARN 资源, 保障迁移过程中不浪费机器资源。
- 5) Hive2 升级到 Hive3, Hive 由 MR 执行引擎, 改为 TEZ 执行引擎。优化 Tez 和 Hive 参数, 启用 Hive3 ACID 特性, 不同业务采用不同的表引擎, 达到能够利用 hive 高效更新、删除数据的效果。
- 6) HBase 用做提供实时计算维表, 导入时使用 BulkLoad 方式, 减少对查询的影响。
- 7) 通过监测每日 80K+ 的作业执行情况, 进行整个集群的 HiveSQL、Clickhouse SQL 优化, 提升作业流性能。

5. **Clickhouse 集群做的优化**

- 1) 去掉 zookeeper 组件, 从磁盘层面保障数据不丢失, 降低了系统的依赖性;
- 2) 全面采用了 BitMap 实现高效精准去重, 性能提升 1-10 倍;
- 3) 利用物化视图特性, 实现流入数据落盘即计算完成;
- 4) 利用聚合表, 对明细数据做预聚合, 提升 10%-50% 性能。

2016/06—2018/12: 北京青果灵动科技股份有限公司

所在部门: 大数据开发组

职位: 高级大数据开发工程师 (兼大数据组组长)

● **职 责:**

1. 负责研发 Hadoop 大数据平台: 秉承“数据化运营”原则, 对数据进行采集、清洗、分类, 然后对数据进行分析和挖掘, 利用分析和挖掘的数据指导运营做出决策, 不断地调整数据挖掘方案以优化数据对人的指导性。
2. 负责调研新的数据发展方向, 丰富大数据平台功能: 调研 TalkingData 等大数据平台的功能, 归

纳和整理数据平台功能和运营方提出的需求，规划大数据平台新的功能。

3. 负责设计 Hadoop 大数据平台架构：根据各个框架市场成熟度，业务契合度，选择适合的框架构建大数据平台，规划业务的扩展性，测试框架在实际数据中的稳定性。
4. 设计大数据平台接口规范：梳理业务需求，整理为技术接口，制定接口规范，供各业务复用。
5. 负责搭建和维护 Hadoop 大数据集群：集群脚本和计划任务优化；集群增删节点，磁盘增加缩减；集群安全性排查；集群 CPU、内存、IO、网络异常排查；Hadoop 文件系统数据均衡维护，Hadoop 文件异常修复；Hbase 数据倾斜维护，Hbase Region 异常修复，Hbase 库的日常备份，恢复等。
6. 负责设计大数据自动化监控系统：自动化监控机器负载情况，监控所有脚本执行状态，监控 Hbase 入库质量，监控 Spark 执行任务状态，监控业务数据是否结果正确。
7. 负责研发公司运营分析后台，对后台数据库进行参数调优，SQL 语句调优，数据库备份和恢复。
8. 负责研发公司官网，内部系统等网站，对网站进行缓存系统调优，数据库索引优化。
9. 负责带动团队解决问题、寻求更高突破：负责带动团队解决业务提出的问题，合理安排团队项目排期，带动团队学习新的技术，优化架构和组件的稳定性和性能。

● 成 果：

1. **FancyData 大数据平台**

FancyData 大数据平台集群规模 100+台服务器，每日 600G+增量数据写入 HBase，已稳定运行 3 年以上。架构底层以 Hbase 和 ElasticSearch 作为数据仓库；上层为离线数据分析框架 Spark SQL 和 Spark Core 出每日/每月报表数据；数据源采用 Fluentd 日志采集工具采集实时写入 Hbase 库。Web 展示平台主要采用 Ruby on Rails 和 Postgresql 数据库。

2. **XData 大数据平台**

XData 大数据平台，专门为 SQL 而设计。架构底层以基于 Hadoop 的 Hive 作为数据仓库；上层以 Presto 作为实时和离线计算工具出每日/每周/每月报表数据；数据源采用 Fluentd 日志采集工具采集实时写入 Hadoop 的 Hive 仓库。Web 展示平台主要采用 PHP 和 MongoDB 数据库。

3. **青果运营分析后台**

青果运营分析后台主要采用 Ruby on Rails 和 Postgresql 数据库，核心技术为数据库 dblink 同步读取技术、分表分区设计。实现原理：通过数据库的 dblink 模块对数据库数据进行同步和关联，通过 ID 取模的值对数据分表，通过月份对数据库分区，实现 Log 库每日高效同步、高效读写、高效统计。

4. **青果官网，青果内部网站**

青果官网包括青果灵动官网（<https://www.fancyguo.com>）和 Fancy3D 引擎官网（<https://www.fancy3d.com>），与青果内部网站采用 LDAP 做统一用户认证，采用 Postgresql、MySQL 数据库，为了网站极速响应，进行缓存优化，数据库索引优化，加入 Redis 做为缓存提升查询效率。

2011/07—2013/05：无锡永中软件有限公司

所在部门：云计算研发部

职位：J2EE 开发工程师

职 责：

1. 负责永中公有云平台（<http://www.yozocloud.cn>）的开发工作。
2. 主要使用的技术：MySQL 主备环境做容灾备份，JCR 做文件库控制文件权限，Spring 框架做业务解耦，Hibernate 框架做数据库关联，ExtJS 做后台展示框架。

自我评价

把自己热爱的事情做到极致。