

简历

个人概况

姓 名：曲佳境 性 别：男
联系电话：18642196696 邮 箱：710552907@qq.com
目标职位：大数据开发工程师 期望地区：北京，上海，杭州等



教育经历

哈尔滨理工大学 软件工程 本科 2016.09-2020.07

大数据相关技能

1. 熟悉 Hadoop 生态圈组件包括 Cloudera Manager, HDFS, Yarn, Hive, HBase, Flume, Sqoop。
2. 熟悉 Spark 开发，对 Spark 调资源有一定的了解。
3. 熟悉 Alluxio 内存为中心的虚拟分布式存储系统，曾用它提高远端读取速度 10-20 倍。
4. 熟悉 AirFlow 任务调度引擎。
5. 熟悉 CM 报警，Yarn 监控，Streaming 监控，以及 HDFS 小文件问题报警。
6. 了解 Kafka 消息队列和 Zookeeper 协调服务，简单了解 Mysql 卡住崩溃排错。
7. 常用的语言 Python, Java, 使用过的语言 Scala 和 Shell。
8. 了解 Zeppelin 应用场景和优势，了解其解释器的配置和使用。
9. 了解大数据可视化，实现实时监控大屏的方案。

附个人博客：<https://shmily-qjj.top/>

实习经历

2019.08.14-2020.3.09 上海合合信息科技发展有限公司（大数据开发）

项目经历

项目一：Spark 自动调参与实时监控

项目描述：基于获取到的 Metrics 监控参数和 Yarn 日志动态调节 Spark 的参数，内存利用率整体提高 2 倍。实现自动调节资源和自动排错。

开发技术：Kafka、Spark、Mysql、Yarn、InfluxDB、Grafana、Airflow

负责内容：

1. 编写获取 spark 内存信息的监控程序并在 spark-submit 时加 javaagent 来调用。
2. 将浪费内存的应用及其 Driver, Executor 使用内存量记录存入 mysql。
3. 针对由资源问题引起的错误，自动应用新参数并重试，若数据量突增，加大内存比率重试。
4. 借助 InfluxDB 和 Grafana 实现实时监控 Spark 应用状态。

项目二：定制版“Sqoop”开发和实现

项目描述：采用 Spark 取代 Sqoop 低效的 MR 来做数据抽取，兼容多种类型表如 Mysql(包括分库分表)，HBase 和 Mongo，实现多种抽取方式全量,增量,拉链表,可视化调度，以及小时跑，日跑和周跑抽取，同时对抽取前后数据量做检测，对异常任务自动报警。

开发技术：Spark、Hadoop、Zookeeper、Hbase、Phoenix、Hive、Mysql、MongoDB、Airflow

负责内容：

1. Spark 代替 Sqoop 大面积提高数据抽取任务效率且抽取逻辑可随时迭代。
2. 自动优化 Spark JDBC 的参数，提高效率。
3. 增量抽取字段类型和范围自动获取，自动获取变更 schema。
4. 实时需求考虑用 Canal 或 Debezium。

项目三：影视大数据分析与访问来源分析

项目描述：根据用户日志进行实时分析，得到用户最喜爱视频类别，根据以上数据实施精准营销及网页设计，并分析网站访问来源，有了数据支撑就能更好地决定在哪个宣传渠道投入更多广告费。

简历

开发技术： Hadoop、Zookeeper、Flume、Kafka、SparkStreaming、HBase、SpringBoot、Echart

负责内容：

- 1.持续调度数据生成器，不断生成日志数据
- 2.Flume 收集多台服务器上的数据并聚合传输到 Kafka
- 3.SparkStreaming 消费 Kafka 的消息，对日志数据做出分析和统计
- 4.将 SparkStreaming 得到的数据写入 HBase
- 5.SpringBoot 读取 Hbase 数据并搭配 Echart 开源可视化库实现数据可视化