


2022.04.13 SSC kafka集群宕机对spx的影响复盘

故障 发生 时间	发生时间	2022.04.13 19:33
	发现时间	2022.04.13 19:36
	恢复时间	2022.04.13 20:22
	持续时间	49min
故障 处理 人	SRE: yingting.miao, zepeng.yao 中间件: baomao.fang	
	SPX: lianjie.chen、huan.chen、chao.chen、 shaofeng.hu 、谢磊、李川	
故障 责任 主体	kafka中间件团队	
故障 报告 来源	Noc监控告警	
故障 类别	一般故障	
故障 描述	SSC kafka集群宕机导致各条业务线作业受影响	

影响评估 TW 反馈包裹出库失败，大概2117单受影响（2022.04.13当日出库总单量为161569，受影响单占比为1.3%）。由于页面显示报错，导致买家需要二次上门取件，影响用户体验和品牌形象。出库失败原因是因为出库接口耗时较长，导致前端访问接口超时（时间为：2022.04.13 19:42~2022.04.13 19:44）。



SPXTW_Jerry Wu


Li Dong (李东) | SPX FPM

@Sarah Pan | SPX BPM 定位到了，是公司的Kafka 集群挂了，导致前端的接口调用报错。只是显示问题，发的这几个单，后台查数据都是正常的，也不需要datafix。 ...

請問有可能知道這個顯示異常的包裹清單嗎？

因為門市人員看到結帳異常的訊息，就會請買家改天再來；但包裹系統上已經被完成取件，款項也沒拿到

20:41



SPXTW_Jerry Wu

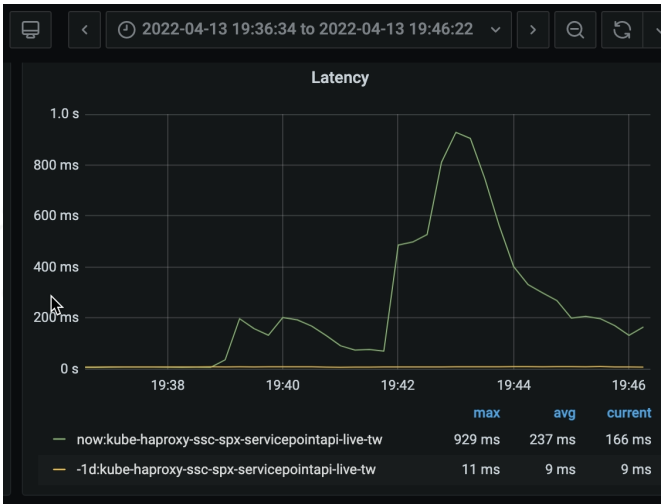
Li Dong (李东) | SPX FPM

我确认下原因。

@Sarah Pan | SPX BPM 因為影響了許多門市，且後續處理相對麻煩（系統已買家取件但實際沒付錢），明早方便提供一下事故報告嗎？以及看後續怎麼預防這個問題發生

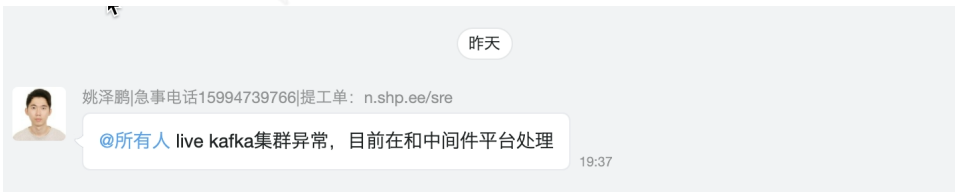
再麻煩了，大家辛苦了～

21:23



处理过程

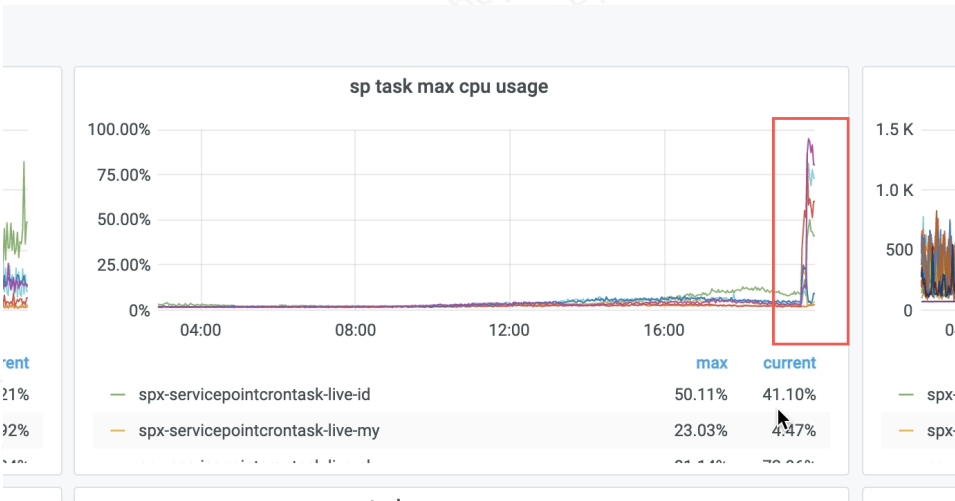
1、04.13 19:37分 Sre在ssc故障大群通告kafka集群故障



2、19:37分, spx 收到rabbitmq堆积告警, Sre开始处理



3、19:37 spx-servicepoint-task cpu保障。于19:43驿站chenhuan做了切换, 不再连kafka, cpu逐步下降, 接口耗时下降;



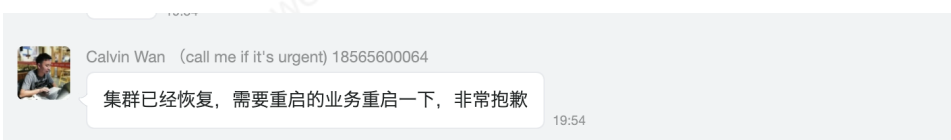
4、19:40 spx-servicepoint-task 拉saturn的同事进行排查。目前初步怀疑是kafka client poll问题。参考

<https://issues.apache.org/jira/browse/KAFKA-10254>

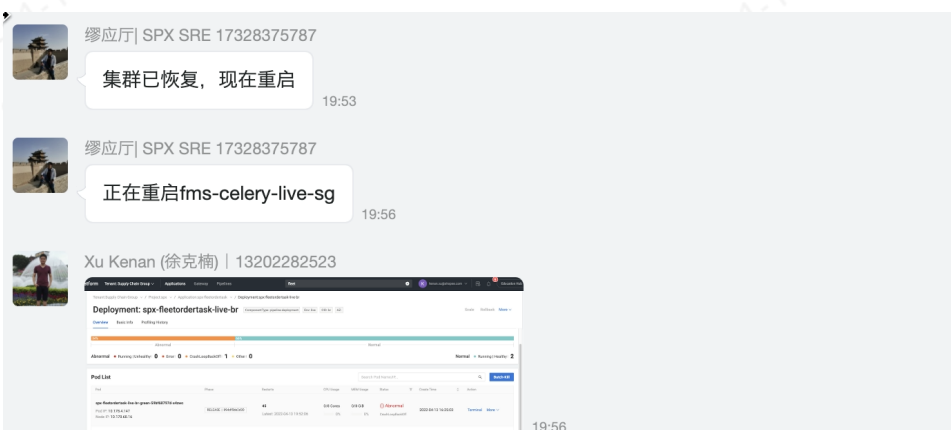
saturn同事会升级kafka client解决此问题。



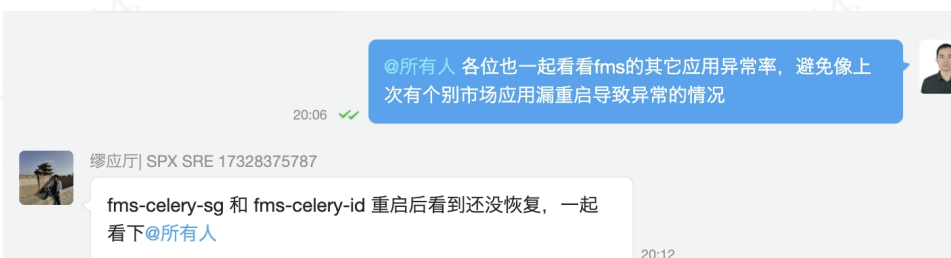
5、19:53 Sre重启kafka节点, 集群恢复



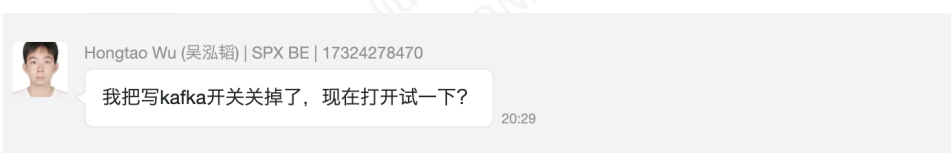
6、19:55 Sre分批重启fms-admin/fms-api/fms-celery等应用



7、重启完毕, 业务恢复, 不过fms-celery-id没消费kafka



8、hongtao确认是由于kafka开关关掉了



9、hongtao重新打开kafka开关, 全部恢复

原因分析	和2022-04-07 SSC kafka集群宕机对spx的影响复盘 一样
改进方案	<p>1、研发侧改进：和2022-04-07 SSC kafka集群宕机对spx的影响复盘 改进策略一样，统一由 Shaofeng Hu 负责</p> <p>2、service point侧改进：service point发送kafka超时优化，kafka故障时，发送快速失败，降低对接口耗时的影响， Huan Chen</p> <p>2.1 驿站取件出库针对网络超时场景优化错误提示信息； CC Jing Huan Chen zhen</p> <p>2.2 同时梳理出和Kafka相关业务场景； Huan Chen</p> <p>3、SRE侧改进：</p> <p>3.1 拆分集群，SPX 拆分计划</p> <p>4、故障上报机制：</p> <p>中间件监控收口： Shaofeng Hu 对接值班PM</p> <p>各团队监控：研发PIC，对接对应业务PM，再依次上升</p>