日志监控总结

## 背景

IBS-日志订阅服务于客户日志查询的场景，支持客户在IBS页面订阅日志，通过API拉取日志，满足客户进行日志比对，确保账单准确的需求。自2023年3月以来，随着客户流量激增，日志量随之激增，日志订阅出现了如下问题：

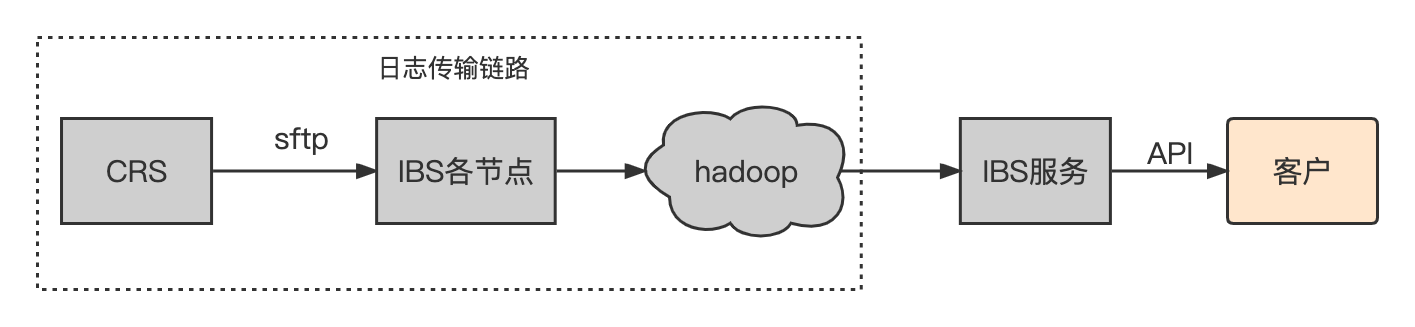
1. **准确性问题**：日志出现丢失的情况，偶发性日志丢失比例10%；
2. **时效性问题**：日志延迟超过最大阈值，出现超过约定的最大阈值3H的比例升高，特别是在大流量客户中更为明显。

以上问题给客户比对日志造成了极大的影响，客户频繁反馈，问题亟待解决。

## 问题分析

经分析，我们得出造成以上问题的原因如下：

1. **日志量激增造成日志传输链路各环节有延迟**，日志传输链路示意图如下：



1. **日志量激增造成原日志中转节点空间吃紧**，单台节点日志存储空间为28T(2台服务器)，所有日志一天为30多T，虽然做了按天删除的策略，但随着日志增加，仍然会出现空间不够，日志无法写入的情况。

## 目标

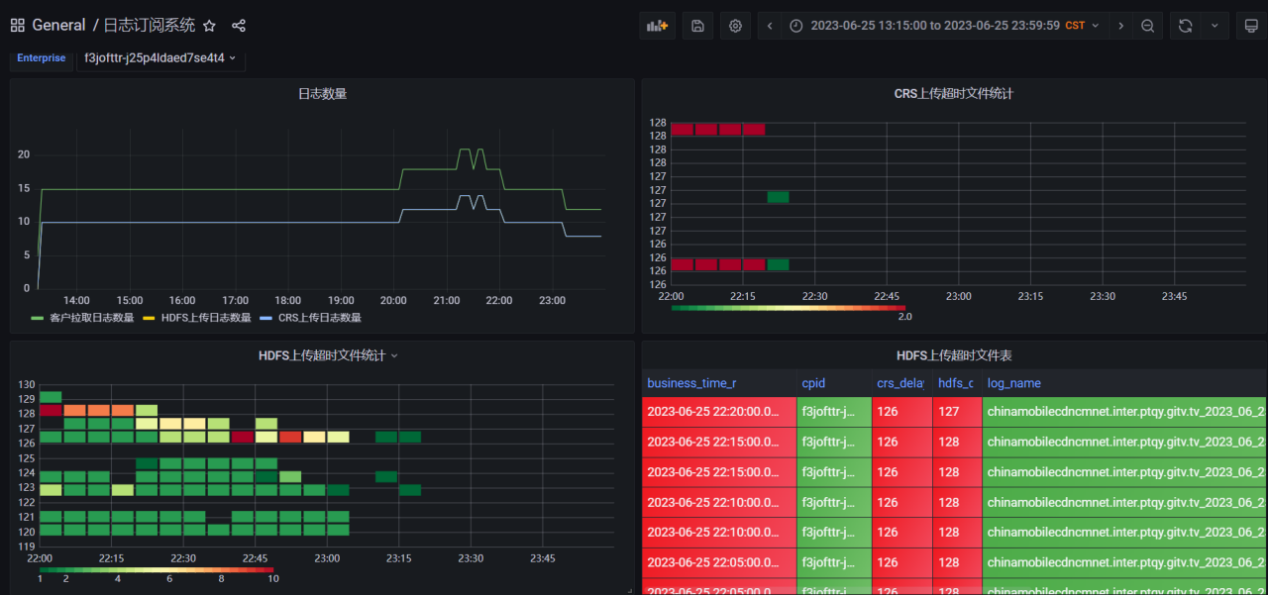
为了彻底解决问题，首先我们需要对当前现状有具体的认知。

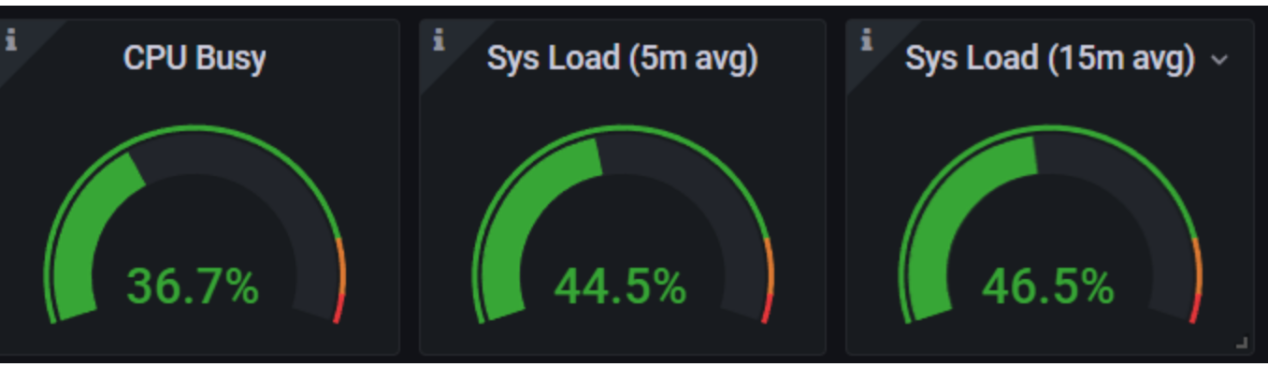
因此本次工作的目标是，**搭建日志监控系统**，对日志传输链路各环节时间消耗、节点健康、网络表现等情况实时监控，**感知问题**，以便于深度分析问题，解决问题。

## 动作和当前成果

整体工作分为两部分：搭建日志监控，展示监控报表。

1. 搭建日志监控系统
   1. 日志采集filebeat -》logstash对日志过滤加工-》ES
2. 展示监控报表
   1. 通过Grafana展示报表，指标有如下几类：
      1. 衡量节点健康：CPU利用率、内存消耗、磁盘使用情况等；
      2. 日志传输链路延时情况、客户日志拉取情况等。
3. 当前成果
   1. 当前在测试环境模拟数据，共展示50多个报表，报表举例截图如下：







## 总结

通过以上工作，实现了以下几点信息的监控：

1. 资源监控：可实时监控到磁盘、内存、CPU等硬件指标。
2. 业务监控：
   1. 可监控crs上传的日志包数量，以及crs的上传延迟；
   2. 可监控上传HDFS的日志包数量，以及上传延迟情况；
   3. 可监控客户的日志下载数量，以及分析客户下载行为；
   4. 异常文件可通过页面进行查找并分析原因。

## 后续规划

1. 自动报警：当监控指标异常时，实时自动报警，及时人工介入处理。