PRD-AIF 技术文档

AIF Spring Cloud Log4X

产品白皮书

声明

- 任何情况下，与本软件产品及其衍生生产品、以及与之相关的全部文文件（包括本文文件及其任何附件中的全部信息）相关的全部知识产权（包括但不不限于版权、商标和技术秘密）皆属于亚信科技（中国）有限公司(“亚信”)。
- 本文件中的信息是保密的，且仅供用户指定的接收人内部使用。未经亚信事先书面同意本文件的任何用户不得对本软件产品和本文件中的信息向任何第三方（包括但不限于除用户指定接收人以外的管理人员、员工和关联公司）进行行行披露、出借、许可、转让、出售分发、传播或进行与本软件产品和本文件相关的任何其他处置，也不不得使该等第三方以任何形式使用本软件产品和本文件中的信息。
- 未经亚信事先书面允许，不得为任何目的、以任何形式或任何方式对本文件进行复制、修改或分发。本文件的任何用户不得更改、移除或损害本文件所使用的任何商标。

- 本文件按“原样”提供，就本文件的正确性、准确性、可靠性或其他方面，亚信并不保证本文件的使用或使用后果。本文件中的全部信息皆可能在没有任何通知的情形下被进一步修改，亚信对本文件中可能出现的任何错误或不准确之处不承担任何责任。
- 在任何情况下，亚信均不对任何因使用本软件产品和本文件中的信息而引起的任何直接损失、间接损失、附带损失、特别损失或惩罚性损害赔偿（包括但不限于获得替代商品或服务、丧失使用权、数据或利利润；或商业中断），责任或侵权（包括过失或其他侵权）承担任何责任，即使亚信事先获知上述损失可能发生。
- 亚信产品可能加载第三方软件，详情请见第三方软件文件中的版权声明。

编 写	PDD 董育兵	编写 时间	
审 核		审核 时间	
文档版本	V00.00.02		

文档修订记录

日期	版本号	描述	著者	审阅者	日期

目录

1 摘要

集中化日志服务中心系统会对散落在各个不同位置的日志数据进行集中实时采集和索引处理，提供搜索、分析、监控和可视化等功能，帮助系统管理员进行线上业务的实时监控、系统异常及时定位、业务数据趋势分析、故障诊断预警。本文将从产品概述、技术架构、主要功能、客户价值、产品优势等几个方面阐述 Log4X 产品。

2 缩略语与术语

术语	描述
服务调用链跟踪	通过服务调用链的方式，把一次请求调用过程中涉及的所有服务节点完整的串联起来，这样就实现了对请求调用路径的监控
traceId	业务办理全局的跟踪 ID，是跟踪的入口点，实际业务中根据需求来决定在哪生成 traceId
spanId	下一层的请求跟踪 ID,这个也是根据实际需求来定义
parentId	上一次请求跟踪 ID，用来将前后的请求串联起来
GC	JAVA 进程垃圾回收
采样率	按一定频次采集数据的比例，如每 100 次数据采集，采样率 1/10，则实际只采集 10 次
CSF	云化服务框架。提供服务编排及服务管控能力的运行框架

3 产品介绍

3.1 产品定位

统一日志实现对系统所涉及范围内的主机、服务器、网络设备、数据库以及各种应用服务系统访问等产生的日志，进行收集、分析和呈现。通过定义日志筛选规则和策略，帮助 IT 管理员从海量日志数据中精确查找关键有用的事件数据，准确定位网络、服务器或业务系统故障并提前识别安全威胁及设备预警，降低系统宕机时间，提升性能、保障系统安全稳定运行。

目标	范围
1、问题追踪： 通过日志不仅可以在应用启动安装配置时发现问题，也可以在程序运行过程中及时发现异常，快速准确定位线上问题。	***1、应用系统日志：*** 含各个业务系统的应用log输出日志，包含Java类应用日志框架采集的日志(log4j, logback) 、输出到日志文件的应用日志、输出到Syslog服务的日志
2、状态监控： 通过实时分析日志，可以监控系统的运行状态，预警系统潜在风险，做到问题早发现、早处理。	
3、异常分类分析： 通过对异常日志数据的分类分析，准确筛选和判定程序运行过程中的异常行为，做到问题及时处理。	***2、中间件日志：*** Redis、Zookeeper、消息队列、Apache、Nginx。
4、请求上下文日志分析： 通过日志数据中的请求标识、用户标识或手机号等特征数据，快速汇总查询关联的日志上下文数据。	***3、数据库日志：*** Oracle、MySQL（数据库事件SQL采集）。
5、主动告警： 定时查询日志处理过程中监控到的异常数据，按照配置的策略主动下发告警，在问题扩大之前介入处理。	***4、业务操作日志：*** 用户对系统中各应用功能进行业务操作的日志（需要单独对日志做定制化输出）。
6、日志中心： 日志中心采集全域业务系统日志数据，提供灵活的统计分析和展示能力，并可支持日志数据服务的能力开放。	***5、应用间接口调用日志：*** 各个应用系统间接口调用日志。

图 3-1 产品定位

3.2功能架构

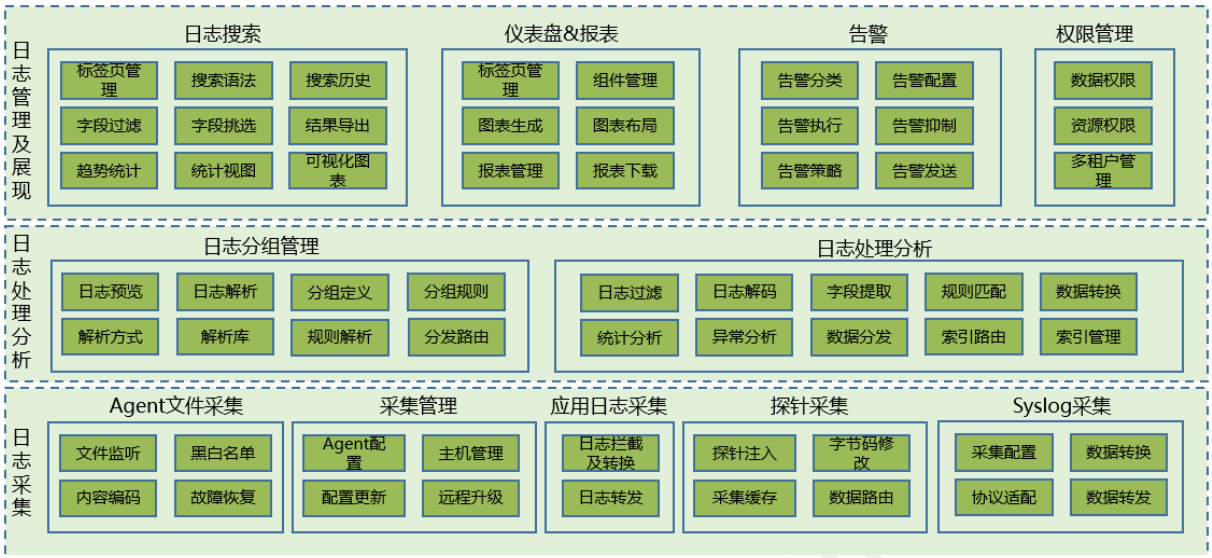

图 3-2 Log4X 功能架构

- 日志搜索：统一日志能力提供日志的搜索功能，可支持多条件组合、多种不同维度进行查询。
- 仪表盘&报表：日志分析结果的展现用于将日志按照安全事件分析场景和业务分析场景，图形化的方式进行交互，用户可根据自身需要快速配置出满足要求的日报、周报、月报。
- 告警：日志告警功能用于各种面向安全事件分析场景和业务系统日志分析的应用交互，对于存在安全隐患或者故障的分析结果，产生告警通知。
- 权限管理：数据权限管理是指用户角色需要严格的数据访问权限的管理控制，支持对用户分组的权限角色设置，统一提供日志来源、告警、已存搜索、仪表盘、趋势图表、字段提取规则、报表等各种资源的权限控制功能。
- 日志分组管理：日志分组是指按照应用系统或租户将日志数据存储在不同的索引或数据表中，达到数据隔离的目的，方便实现日志数据的访问权限控制。

- 日志采集：统一日志能力需要满足收集主机设备、主流操作系统及应用日志信息的需求。对于应用系统，需要收集用户的操作行为日志、系统后台操作日志及异常日志、提供一种统一的日志数据标准并在收集过程中需要注意关键信息的脱敏处理。

3.3技术架构

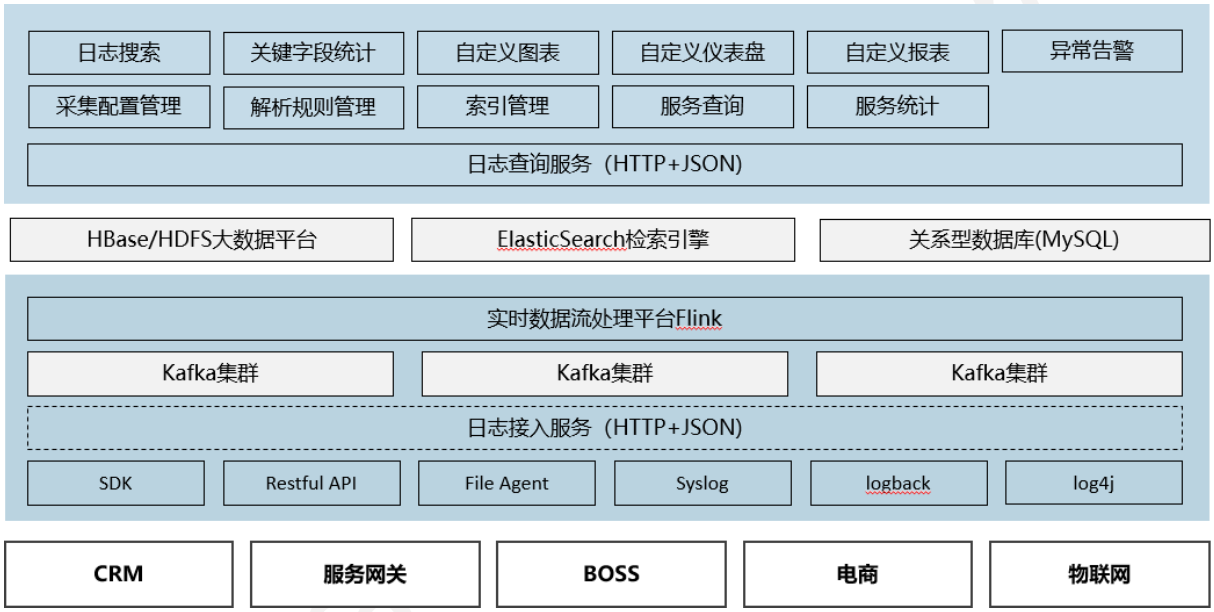

图 3-3 Log4X 技术架构

- 日志采集：统一日志网关服务，支持按SDK/API方式采集和基于Restful接口服务方式采集；
- 日志接入服务网关：日志接入服务提供标准Syslog和HTTP+json两种方式的数据接入，可对接主机日志和服务化日志数据采集。
- Flink：支持高吞吐、低延迟、高性能的流处理，兼容hadoop。
- 日志查询服务网关：日志查询服务提供标准的Restful协议数据输出能力，可通过灵活的搜索表达式检索数据。
- 接口认证：日志服务接口均提供身份认证及数据校验能力，必须经过授权和认证才可接入系统

4 产品价值

Log4X 实现对主机、服务器、中间件、数据库以及各种应用服务系统访问过程中产生的日志，进行统一收集、多维分析和灵活呈现，从海量日志中精确查找事件数据。支持前后台业务全链路跟踪和关键行为回溯分析，一键查找业务调用故障点。日志平台提供能力给运维人员进行快速故障分析定位，提供能力给用户进行异常或业务统计分析，提供数据的开放能力给其它数据平台。

- **无侵入性**

无侵入采集，对业务透明，开发人员不需要关注日志如何采集。无侵入埋点，不对源码进行修改和埋点代码添加；

- **低消耗**

日志系统对在线业务系统的影响足够小，通过异步线程，批量进行日志采集和记录。

- **灵活的应用策略**

可以随时变更收集数据的范围和粒度，如地市、时段等。应用服务器中日志文件的留存时间，日志数据采集的频率等。

- **时效性**

从数据的产生和收集，到数据计算和处理，再到最终存储，有极高的实时性。

- **决策支持**

通过 Storm、Hadoop 等大数据技术实时或全量分析收集的日志数据为决策提供数据支持。

5 产品部署

5.1 部署框架

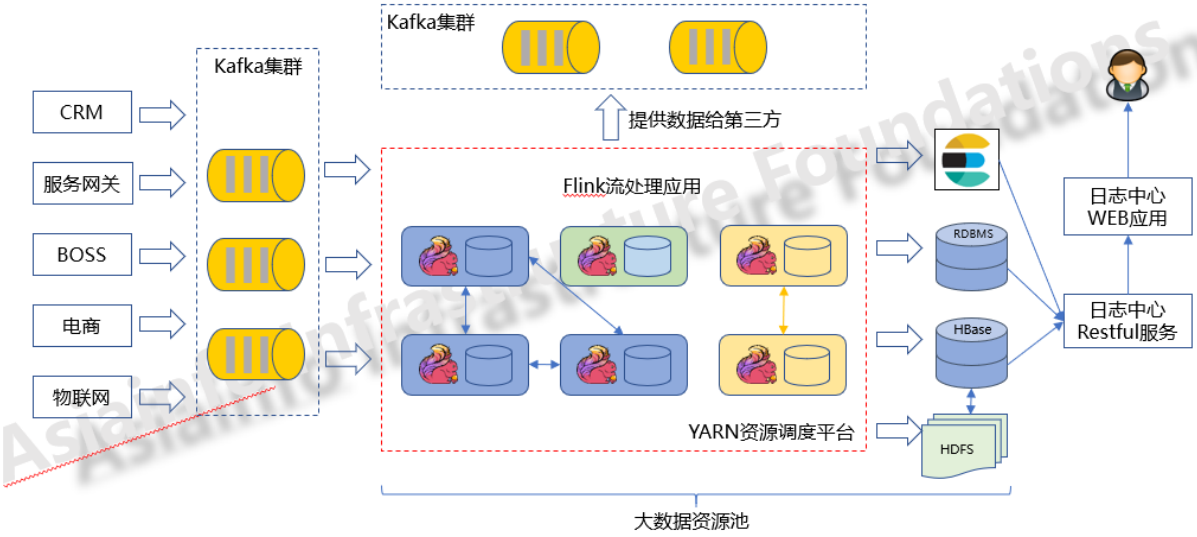

图 5-1 部署框架图

- 日志采集模块：Log4x的客户端，通过Agent和Client将采集部分代码通过字节码增强或者植入代码到应用服务系统中，完成数据的采集，并发送到kafka；
- 日志分析模块：以Flink作为实时流处理的核心系统，完成Trace日志的处理、统计和预警功能，并将数据存储于Hbase和RDBMS中；
- 日志可视化模块：日志应用服务器完成日志统计结果的查询和搜索，通过CSF框架提供数据供web端展现。图中WEB Server包括了csfproxy，csfserver和web三个部分；
- 数据查询服务开放：通过Restful接口，将日志中心原始数据或沉淀后的数据开放给第三方平台。可利用日志中心实时统计分析能力，将日志数据进行分析处理后，将统计结果开放给第三方平台；

5.2运行环境要求

- JDK1.8或 以上版本

5.3第三方软件

软件	版本号	描述
Kafka	0.11	日志数据缓存
Zookeeper	3.4	分布式协调服务
Flink	1.7	流处理平台
ElasticSearch	5.6	全文检索引擎
HBase	1.2.6	服务调用明细数据
Hadoop	2.6	分面式文件系统
MySQL	5.6	基础数据
nginx	1.1.5	服务网关代理

6 产品案例

6.1北京联通公司

建设背景：北京中台系统架构按多层、多中心化设计，分布式部署，并且随着系统能力的提升及业务规模的急剧扩大，应用系统产生了大量分布的日志数据，运维困难。

总体架构：

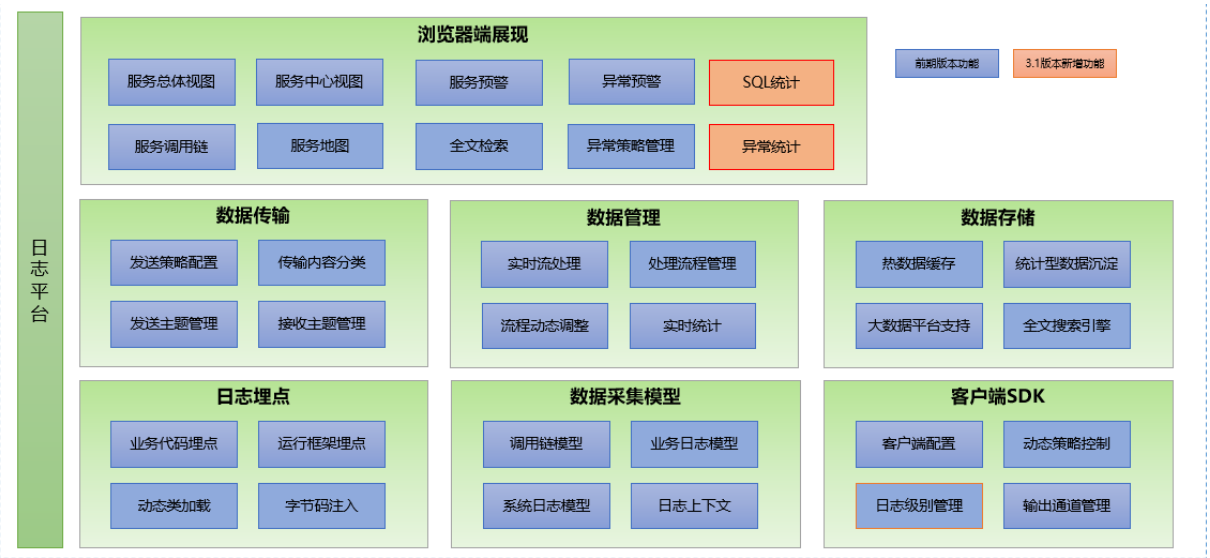

图 6-1 北京联通 Log4X 系统架构

建设效果：Log4X 采用流处理技术分析实时汇总数据，采用大数据技术进行日志数据的存储和分析。快速数据分析，日志能够快速导出及分析，能够快速的定位系统的异常，及时应对，提升系统的稳定性。具体效果如下：

- 调用链信息展现可视化，便于进行相关性能损耗、异常分析等；
- Google Dapper 模型实现对 RPC 调用的链路监控；
- LOG4X 系统跟踪功能集成到 CRM 各个基础框架功能组件中；

7 联系我们

AIF 论坛

你可以在论坛搜索和发言提问，我们将尽可能回答各类问题，不能灌水哦！

点这里直达 AIF 论坛（内网访问）。

AIF 常见问题

你可以在这里找到常见的问题，帮助你快速 get 到 AIF 各种安装、部署、发布等等问题。

点这里直达 AIF 常见问题（内网访问）。

AIF 服务邮箱

如果你有任何建议或者意见，请点击下面链接发送邮件，我们将尽可能回答你的问题。

aif@asiainfo.com