

应用系统

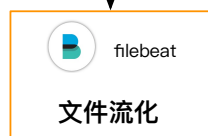
应用系统进行日志打点，
通过LogSDK以Http方式上传

互联网接入



日志接入层

通过Nginx接收日志请求，
写入磁盘日志文件



日志归集层

实时汇集日志文件，
统一存储到kafka集群中



日志应用层

读取归集层日志流，
进行规范化格式转换后存储
到应用层日志存储



注：建议应用层和归集层使用
不同的kafka集群



Kylin

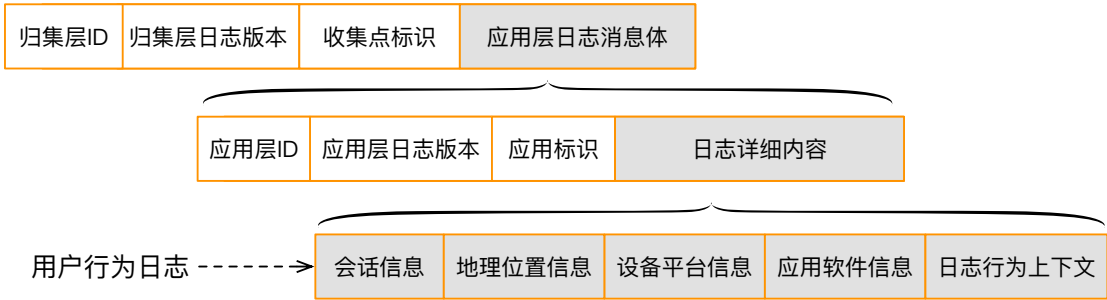


HDFS



HIVE

日志格式规范:



归集层日志

域	字段内容
归集层ID	全局唯一的字符串,标识单条日志消息
归集层日志版本	版本字符串
收集点标识	节点标识符 + 应用标识符 + 格式标识符
日志消息体	内容格式信息 + 内容体（部分HTTP头 + HTTP请求体）

应用层日志

域	字段内容
应用层ID	日志ID: logId, 全局唯一,标识单条日志
应用层日志版本	版本号: logVersion
应用标识	应用ID: appId, 唯一标识某个产品线内的某个具体应用
日志详细内容	和appId相关的日志详细内容

用户行为日志内容

域	字段内容
会话信息	会话标识: sessionId 账号标识: accountId 用户ID: userId
地理位置信息	设备IP地址: ip
设备平台信息	终端设备型号: productModel
应用软件信息	app系列号: appSeries app版本号: appVersion
日志行为上下文	日志类型: logType 日志发生时间: happendTime 日志接收时间: receiveTime 其他logType特定的字段 logCount pageID controllerId pageContext controllerContext

日志存储规范:



日志文件

日志文件: 名称映射到产品线标识,txt格式的日志轮滚文件
保存时长: 遵循日志文件备份策略



Kafka

日志归集存储

topic名称: 映射到产品线标识
消息体: 单行日志文本的UTF-8字节数组
保存时长: N天,视具体情况定



Kafka

应用日志存储

topic名称: 映射到产品线标识和应用标识的组合
消息体: 规范化日志记录的序列化字节数组
目前采用JSON格式记录的UTF-8字节数组
保存时长: N天,视具体情况定



HDFS

目录结构: 根目录/产品线标识+应用标识/时间分区
文件: 存储应用层日志记录的序列化结果
保存时长: 遵循日志文件备份策略

