**日志采集需求文档**

V1.0.0

王智强 @ Enlink 数据中心

**目录**

[1. 概述 3](#_Toc30293)

[1.1. 说明 3](#_Toc26681)

[2. 日志文件要求及建议 4](#_Toc21642)

[2.1. 文件路径及命名 4](#_Toc13451)

[2.2. 日志内容排版 4](#_Toc9431)

[2.3. 日志字段 4](#_Toc25978)

[2.4. 其它情况 5](#_Toc26543)

[2.5. 交付清单 5](#_Toc27305)

# 概述

## 说明

本文档描述Filebeat对日志原始数据清洗所需要的命名，排版，字段格式等要求及建议，方便日志产生者对日志的处理以及减少不必要的沟通。

# 日志文件要求及建议

## 文件路径及命名

Filebeat支持在多条路径下扫描日志，且为多路径同时扫描跟配置路径的顺序无关，可根据需要把日志源文件放到不同的文件夹里，相同路径下的日志文件命名则要求命名统一，例如：<name>-<yyyy-MM-dd>.log，尽量避免日志数据都放在同一文件里。路径也支持多重星号目录，例如：/var/log/\*/<name>-\*.log

## 日志内容排版

日志内容的排版要求整洁有序，格式统一化规范化，拒绝内容里出现眼花缭乱的符号，例如：“====”，“++++”，“——”等。日志内容要求尽可能完整，每条日志都完整的描述了：什么时间什么场景下发生了什么事以及哪些可能原因及解决提示等。

日志各个字段之间应用字符分（“ ”，“|”等）隔开以方便后续清洗，有关联的字段应放在一起，例如表2-1：

表 2-1 日志排版示例

|  |
| --- |
| 时间戳|日志等级|账号|用户组|会话ID|源IP|主机名|流量信息|请求方法|HTTP版本号|URL|状态码|响应时间|ContentType|UserAgent|Refer(URL) |

若有日志因系统等原因产生不同类型的字段及日志内容排版无法统一则需将统一的字段放日志前面，其他不同字段放最后，例如表2-2：

表 2-2 日志排版示例

|  |
| --- |
| 时间戳|日志级别|账号|用户组|会话ID|源IP|操作|结果|失败原因|认证服务器  时间戳|日志级别|账号|用户组|会话ID|源IP|操作|结果  时间戳|日志级别|账号|用户组|会话ID|源IP|操作|结果|异常情况 |

## 日志字段

日志里的同一字段的格式要统一，大小写统一，字段前后不应出现除分隔符之外的无关字符。字段值假如为空时，分隔符为“|”时直接为空即可，分隔符为“ ”时候则需用其他约定好的字符代替。

字段的内容因尽量具体，比如NC资源不足，究竟具体指什么资源不足，是否可以通过程序直接指明。与该日志无关的字段应舍去。通用字段的格式应与惯用规范一致，例如表2-3：

表 2-3 日志字段

|  |
| --- |
| 时间戳：YYYY/MM/DD-HH:MM:SS，YYYY/MM/DD HH:MM:SS等  日志级别：fatal,error,warn,info,debug  SessionID：长度32位 |

## 其它情况

通常情况下日志文件里不因该有其它不符合约定格式的日志，否则会导致该日志索引会有其它杂乱数据，如果因特殊情况日志文件中有大量其它与该日志无关的日志则需要对其进行过滤清洗，及通过正则表达式把无关日志过滤掉，推荐做法为在该日志中加入其它无关日志没有的标识字段，例如表2-4通过添加特殊标识字段（“ENSERVER”）进行区分：

表 2-4 日志示例

|  |
| --- |
| 2019/01/02 17:27:43 [emerg] 2128#0: user quit: | name[] | ip[192.168.100.65:64538]  2019/01/02 17:27:43 [emerg] 2128#0: user quit: | name[] | ip[192.168.100.65:64538]  2019/01/02 17:27:43 [emerg] 2128#0: ENSERVER:|LOGOUT|xiaoming|192.168.100.65:64538|||  2019/01/02 17:27:43 [emerg] 2128#0: ENSERVER:|LOGOUT|xiaohong|192.168.100.65:64538|||  2019/01/02 17:27:43 [debug] 2128#0: \*1 reusable connection: 0 |

如果因特殊情况一条日志记录为多行显示，则第一行开头需要有其它行没有的特殊标识字段或者格式，例如表2-5：

表 2-5 日志示例

|  |
| --- |
| [2016-05-25 12:39:04,744][DEBUG][action.bulk] [Set][3] failed to execute bulk item (index) index  MapperParsingException  [Field name [events.created] cannot contain ] at org.elasticsearch.index.mapper.object.ObjectMapper$TypeParser.parseProperties(ObjectMapper.java:273)  [2016-05-26 15:39:04,744][DEBUG][action.bulk] [Set][3] failed to execute bulk item (index) index  MapperParsingException  [Field name [events.created] cannot contain ] at org.elasticsearch.index.mapper.object.ObjectMapper$TypeParser.parseProperties(ObjectMapper.java:273) |

## 交付清单

为了方便日志的高效采集及日后维护工作，除了上述格式要求之外还需要以下信息：

1. 日志描述的信息以及由什么系统或模块产生
2. 日志对应的系统版本
3. 该日志所在环境及路径
4. 日志数据源文件
5. 日志信息里每个字段的含义和取值范围
6. 系统产生该日志的频率