归集层消息ID算法（msgId）：

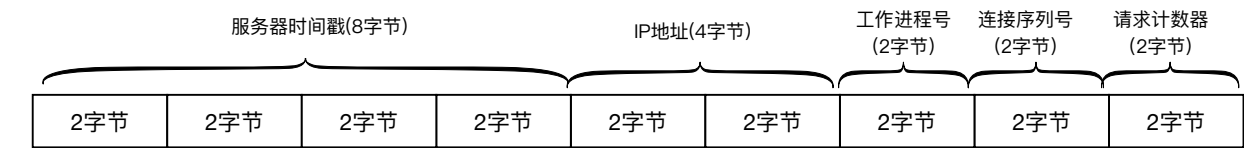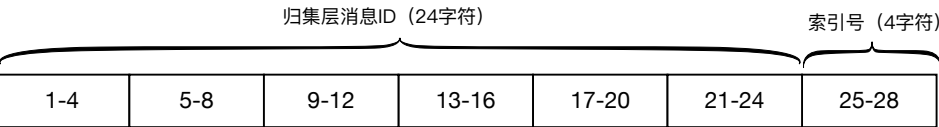| 服务器时间戳(8字节)				IP地址(4字节)		工作进程号(2字节)	连接序列号(2字节)	请求计数器(2字节)
2字节	2字节	2字节	2字节	2字节	2字节	2字节	2字节	2字节

18字节的Base64编码字符串（24位）：
1. 服务器时间戳： 取ngx.now()*1000得到毫秒数，8位字节
2. ip地址： 取ngx.var.server_addr，4字节
3. 工作进程号：ngx.work.pid() % 65536，2字节
4. 连接序列号：ngx.var.connection % 65536，2字节
5. 请求计数器：ngx.var.connecton_requests % 65536，2字节

应用层消息体ID算法（logId）：

归集层消息ID（24字符）						索引号（4字符）
1-4	5-8	9-12	13-16	17-20	21-24	25-28

28位定长字符串：
1. 归集层消息ID： 24位字符串
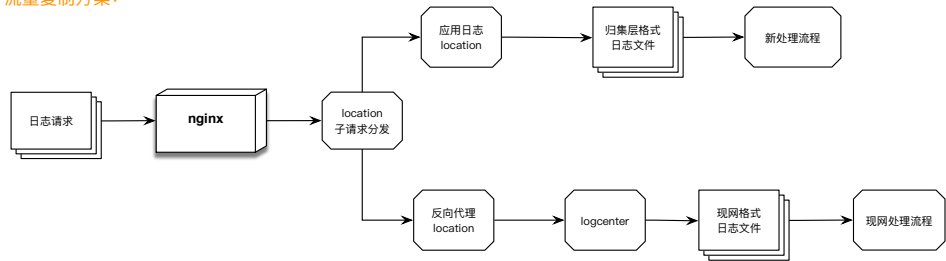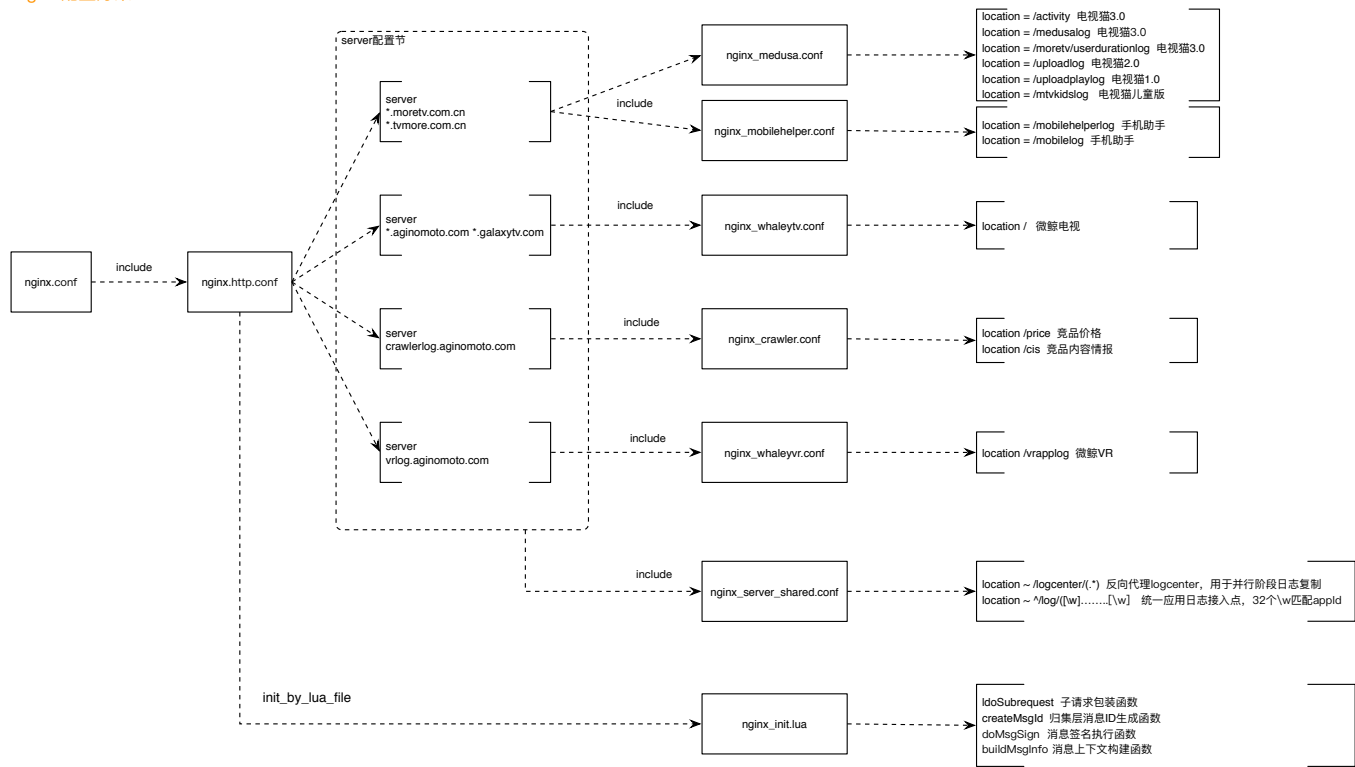2.索引号： 一个归集层消息派生最多65536个应用层日志消息，每个消息分配一个从0开始的2字节索引号，2字节转为4位十六进制字符串

应用ID算法（appId）：
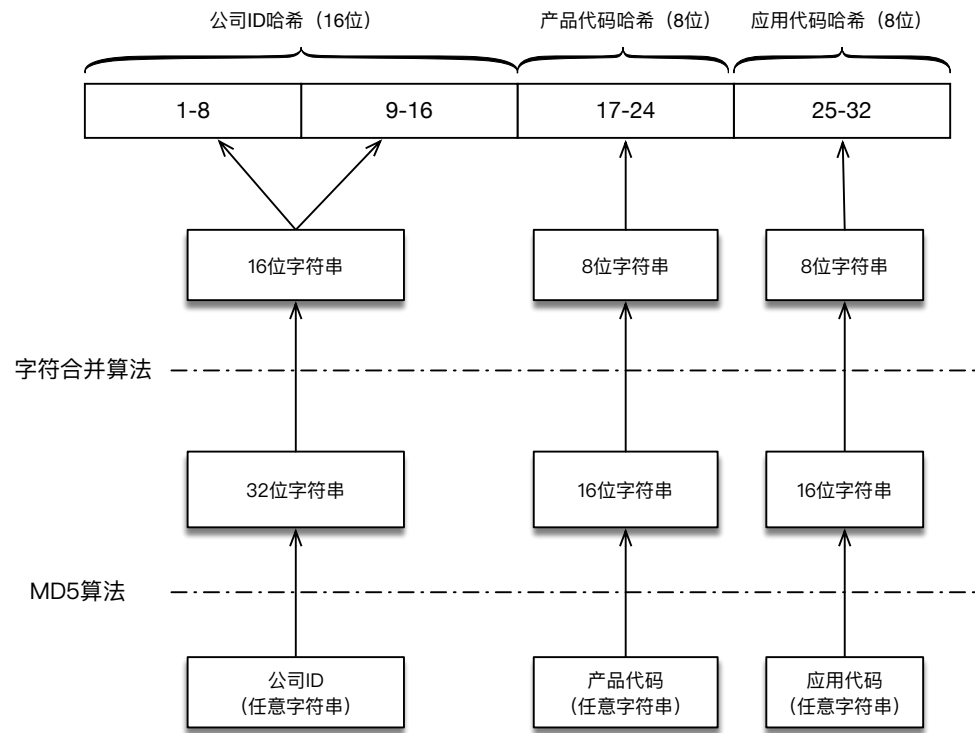
字符串合并算法：

将一个16位0–9a–f的字符串合并为8位0–9a–Z的字符串
1. 定义字符空间A（0–9a–v），从0开始依次编号
2. 16位字符串的每一位映射到编号0–15，即0–9a–f映射到0–15
3. 16位字符串的前后8位编号按位相加再加1，得到8个1–31的编号
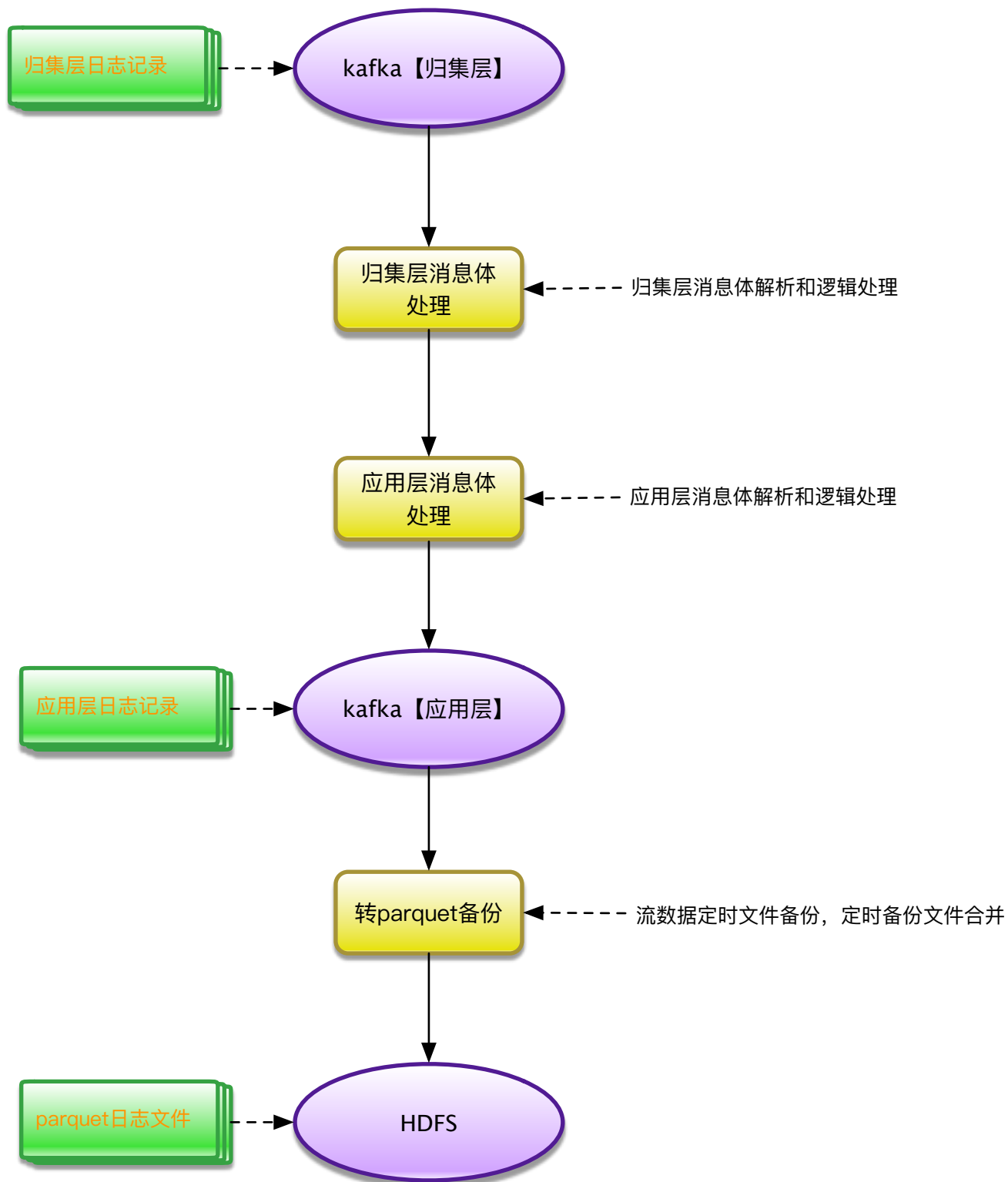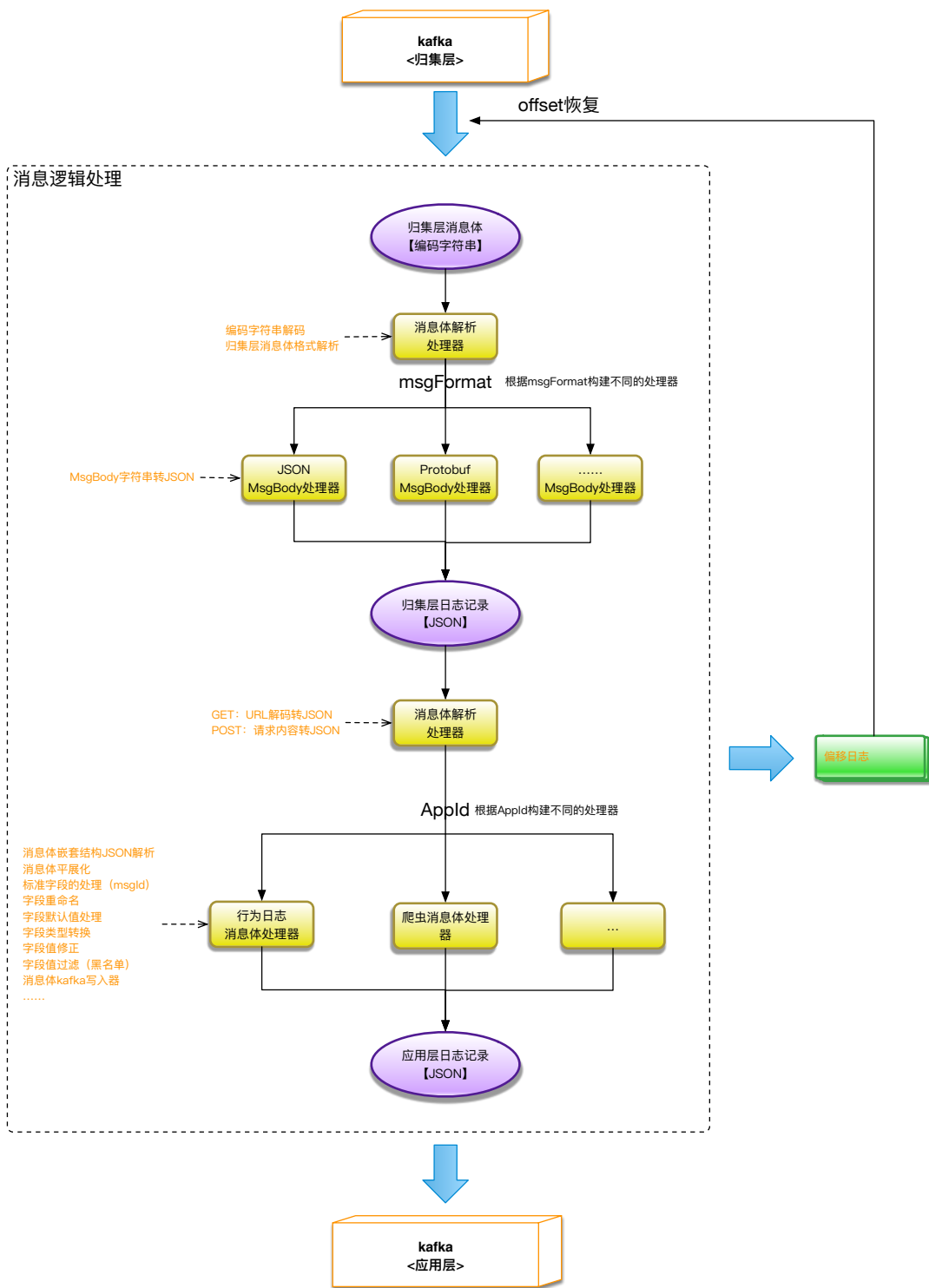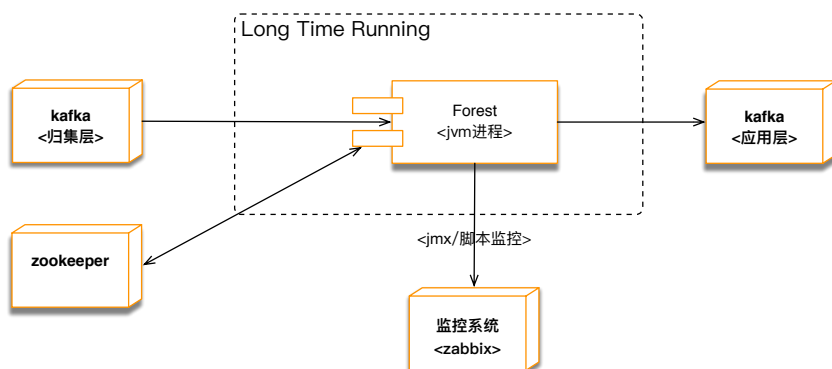4. 以8个1–31的编号为下标，则字符串空间A中取对应的字符组成一个8位字符串

流量复制方案：

日志请求 → nginx → location 子请求分发
- → 应用日志 location → 归集层格式 日志文件 → 新处理流程
- → 反向代理 location → logcenter → 现网格式 日志文件 → 现网处理流程

nginx配置方案：

nginx.conf --include--> nginx.http.conf

server配置节

- server *.moretv.com.cn *.tvmore.com.cn --include-->
  - nginx_medusa.conf →
    - location = /activity 电视猫3.0
    - location = /medusalog 电视猫3.0
    - location = /moretv/userdurationlog 电视猫3.0
    - location = /uploadlog 电视猫2.0
    - location = /uploadplaylog 电视猫1.0
    - location = /mtvkidslog 电视猫儿童版
  - nginx_mobilehelper.conf →
    - location = /mobilehelperlog 手机助手
    - location = /mobilelog 手机助手
- server *.aginomoto.com *.galaxytv.com --include--> nginx_whaleytv.conf → location / 微鲸电视
- server crawlerlog.aginomoto.com --include--> nginx_crawler.conf →
  - location /price 竞品价格
  - location /cis 竞品内容情报
- server vrlog.aginomoto.com --include--> nginx_whaleyvr.conf → location /vrapplog 微鲸VR
- --include--> nginx_server_shared.conf →
  - location ~ /logcenter/(.*) 反向代理logcenter，用于并行阶段日志复制
  - location ~ ^/log/([\w].......[\w]) 统一应用日志接入点，32个\w匹配appId

init_by_lua_file → nginx_init.lua →
- ldoSubrequest 子请求包装函数
- createMsgId 归集层消息ID生成函数
- doMsgSign 消息签名执行函数
- buildMsgInfo 消息上下文构建函数

接入层规划：

appId	公司Id	产品代码	应用代码	应用说明	目前接入点	规划接入点
boikgpokn78sb95ktmsc1bnkk1f477ap	whaley	medusa	main3.0	电视猫3.0版本主程序	domain： log.moretv.com.cn log.tvmore.com.cn log2.tvmore.com.cn log3.tvmore.com.cn 16机器外网IP location: /activity /medusalog /moretv/userdurationlog	domain： *.moretv.com.cn *.tvmore.com.cn location: /log/boikgpokn78sb95ktmsc1bnkk1f477ap
boikgpokn78sb95ktmsc1bnken8tuboa	whaley	medusa	main2.0	电视猫2.0版本主程序	domain： log.moretv.com.cn log.tvmore.com.cn log2.tvmore.com.cn log3.tvmore.com.cn location: /uploadlog	domain： *.moretv.com.cn *.tvmore.com.cn location: /log/boikgpokn78sb95ktmsc1bnken8tuboa
boikgpokn78sb95ktmsc1bnkbe9pbhgu	whaley	medusa	main1.0	电视猫1.0版本主程序	domain： log.moretv.com.cn log.tvmore.com.cn log2.tvmore.com.cn log3.tvmore.com.cn location: /uploadplaylog	domain： *.moretv.com.cn *.tvmore.com.cn location: /log/boikgpokn78sb95ktmsc1bnkbe9pbhgu
boikgpokn78sb95ktmsc1bnkfipphck1	whaley	medusa	kids	电视猫少儿版	domain： log.moretv.com.cn log.tvmore.com.cn log2.tvmore.com.cn log3.tvmore.com.cn location: /mtvkidslog	domain： *.moretv.com.cn *.tvmore.com.cn location: /log/boikgpokn78sb95ktmsc1bnkfipphck1
boikgpokn78sb95kjtihcg268dc5mlsr	whaley	mobilehelper	main	电视猫收集助手	domain： log.moretv.com.cn log.tvmore.com.cn log2.tvmore.com.cn log3.tvmore.com.cn location: /mobilehelperlog /mobilelog	domain： *.moretv.com.cn *.tvmore.com.cn location: /log/boikgpokn78sb95kjtihcg268dc5mlsr
boikgpokn78sb95kjhfrendo8dc5mlsr	whaley	whaleytv	main	微鲸电视主程序	domain： applog.aginomoto.com appstorelog.aginomoto.com appstorelog.galaxytv.cn crashlog.aginomoto.com tvlog.aginomoto.com tvlog.galaxytv.cn videolog.aginomoto.com location: /crashlog /applog /sysapplog /appstorelog /accountlog /activitylog/ /launcher/ /operationlog/ /paylog/ /romlog/ /tvlog/ /upgradelog/ /userinforlog/ /adlog/ /operationlog/ /pagelog/ /videolog/ /playqoslog	domain： *.aginomoto.com *.galaxytv.cn location: /log/boikgpokn78sb95kjhfrendo8dc5mlsr
boikgpokn78sb95kbqei6cc98dc5mlsr	whaley	whaleyvr	main	VR主程序	domain: vrlog.aginomoto.com location: /vrapplog	domain: vrlog.aginomoto.com location: /log/boikgpokn78sb95kbqei6cc98dc5mlsr
boikgpokn78sb95kk1s3bhmt.jqosocdj	whaley	crawler	price	爬虫竞品价格	domain: crawlerlog.aginomoto.com location: /price	domain: crawlerlog.aginomoto.com location: /log/boikgpokn78sb95kk1s3bhmt.jqosocdj
boikgpokn78sb95kk1s3bhmtich.jhhm8	whaley	crawler	cis	爬虫竞品内容情报	domain: crawlerlog.aginomoto.com location: 暂时未定	domain: crawlerlog.aginomoto.com location: /log/boikgpokn78sb95kk1s3bhmtich.jhhm8

```
┌──────────────┐                    ╭─────────────────────╮
│  归集层日志记录  │ - - - - - - →    │    kafka【归集层】    │
└──────────────┘                    ╰─────────────────────╯
                                              │
                                              ↓
                                    ┌──────────────┐
                                    │  归集层消息体   │ ← - - - - - 归集层消息体解析和逻辑处理
                                    │    处理      │
                                    └──────────────┘
                                              │
                                              ↓
                                    ┌──────────────┐
                                    │  应用层消息体   │ ← - - - - - 应用层消息体解析和逻辑处理
                                    │    处理      │
                                    └──────────────┘
                                              │
                                              ↓
┌──────────────┐                    ╭─────────────────────╮
│  应用层日志记录  │ - - - - - - →    │    kafka【应用层】    │
└──────────────┘                    ╰─────────────────────╯
                                              │
                                              ↓
                                    ┌──────────────┐
                                    │  转parquet备份  │ ← - - - - - 流数据定时文件备份，定时备份文件合并
                                    └──────────────┘
                                              │
                                              ↓
┌──────────────┐                    ╭─────────────────────╮
│ parquet日志文件 │ - - - - - - →    │         HDFS        │
└──────────────┘                    ╰─────────────────────╯
```

Forest逻辑处理流程

kafka
<归集层>

offset恢复

消息逻辑处理

归集层消息体
【编码字符串】

编码字符串解码
归集层消息体格式解析

消息体解析
处理器

msgFormat　根据msgFormat构建不同的处理器

MsgBody字符串转JSON

| JSON MsgBody处理器 | Protobuf MsgBody处理器 | ...... MsgBody处理器 |

归集层日志记录
【JSON】

GET：URL解码转JSON
POST：请求内容转JSON

消息体解析
处理器

偏移日志

AppId　根据AppId构建不同的处理器

消息体嵌套结构JSON解析
消息体平展化
标准字段的处理（msgId）
字段重命名
字段默认值处理
字段类型转换
字段值修正
字段值过滤（黑名单）
消息体kafka写入器
......

| 行为日志 消息体处理器 | 爬虫消息体处理器 | ... |

应用层日志记录
【JSON】

kafka
<应用层>

Forest组件结构

Long Time Running

kafka
<归集层>

Forest
<jvm进程>

kafka
<应用层>

zookeeper

<jmx/脚本监控>

监控系统

Backupper逻辑处理流程

kafka
<应用层>

offset恢复

备份到临时文件

应用层日志记录
【JSON】

实时微批备份

临时文件

偏移日志

临时文件合并

定时合并临时文件

parquet文件

HDFS

Backupper组件结构

Long Time Running

kafka
<应用层>

backupper-front

临时文件

Scheduled Running

backupper-combinner

HDFS

zookeeper

<监控>

<监控>

监控系统