



黑客派

# 基于Flume的电力日志收集系统架构设计初探

作者: [flhuoshan](#)

原文链接: <https://hacpai.com/article/1428976058279>

来源网站: 黑客派

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)



an>, <span>MemoryChannel</span>等不同类型的, 其作用可能类似于<span>Java</span>中<span>PipeLine</span>), <span>Sink</span>组件从<span>Channel</span>中获取数据一个<span>Sink</span>只能从一个<span>Channel</span>中获取数据, 而一个<span>Source</span>的数据可以存入多个<span>Channel</span>, 实际上是同时复制了多份数据存入<span>Channel</span>中。<span></span> </p>

<p> <span></span> <span></span> </p>

<p> <i>图<span>2-2clientAgent</span></i> </p>

<p> 实际上, 包括源数据接入层和中心调度服务器层, 无论是<span>clientAgent</span>还是<span>FlumeControllerAgent</span>, 主要都是由<span>Source</span>, <span>Channel</span>及<span>Sink</span>三个组件构成。源数据接入层的<span>Sink</span>组件获取到数据后可以通过配置和编码实现定制化的功能, 将数据<span>sink</span>到中心调度服务器的<span>Source</span>组件。基于系统可靠性和稳定性的考虑, 需要在源数据接入层和中心调度服务器层之间实现<span>loadBalance</span>和重试机制。实现后, 需要处理的各<span>task</span>均衡的载在各中心调度服务器; 同时, 当某台中心调度服务器无法服务时, 由<span>clientAgent</span>通过<span>Sink</span>组件将数据发送到另一台中心调度服务器。源数据服务层支持线性扩展, 论上, 只需要在客户机上安装<span>agent</span>程序并进行相应的配置操作即可将新加的服务日志加入日志处理队列。<span></span> </p>

<script async src="https://pagead2.googlesyndication.com/pagead/js/adsbygoogle.js"></script>

<!-- 黑客派PC帖子内嵌-展示 -->

<ins class="adsbygoogle" style="display:block" data-ad-client="ca-pub-5357405790190342" data-ad-slot="8316640078" data-ad-format="auto" data-full-width-responsive="true"></ins>

<script>

(adsbygoogle = window.adsbygoogle || []).push({});

</script>

<p> <b><i><span>2.2.2</span></i></b> <b><i>中心调度服务器</i></b> <b><i><span></span></i></b> </p>

<p> 中心调度服务器层设计为服务器集群, 在性能达到瓶颈时, 同样可以采用线性扩展的方式对集进行扩容, 而这只需要变更配置文件及少许代码的修改而已。中心调度服务器的主要工作是接收来自据源接入层的数据, 根据路由规则将数据送到不同的存储服务器。<span></span> </p>

<p> <span></span> <span></span> </p>

<p> <i>图<span>2-3</span></i>: <span>FlumeControllerAgent</span></i> </p>

<p> 如图所示, 数据从<span>clientAgent</span>传递到<span>FlumeControllerAgent</span>后, <span>Source1</span>、<span>Source2</span>、<span>Source3</span>中的数据会汇聚到<span>Channel</span>中, 作为<span>Sinkhkd</span>, <span>sinkkafka,sinkbyass</span>的数据来源。当然, 一个<span>FlumeControllerAgent</span>中也可以有多个<span>Channel</span>, 用以实现不同数据的隔离。<span></span> </p>

<p> <b><i><span>2.2.3</span></i></b> <b><i>存储服务器</i></b> <b><i><span></span></i></b> </p>

<p> 存储服务器, 设计分为<span>HdfsServer</span>、<span>KafkaServer</span>、<span>ByPassServer</span>。<span>HdfsServer</span>分布式存储所有日志, 可供<span>Hadoop</span>进行离线分析使用; <span>KafkaServer</span>可以存储某段时间的日志(可配置), 同为<span>Storm</span>框架提供实时的流数据, 供实时在线分析; <span>ByPassServer</span>的本质也是一台<span>Agent</span>, 负责为其他的框架提供实时的日志流, 可以通过配置和少编码将日志流<span>sink</span>到其他框架, 供分析使用。<span></span> </p>

<p> <b>三、目前可能存在的问题</b> </p>

<p> 目前<span>FLume\_NG</span>框架中, 主要提供了下面的几种<span>Channel:</span> <p>

<table>

Channel	说明
Memory Channel	数据存储存储在内存中
JDBC Channel	数据存储存储在持久化存储中，当前
File Channel	数据存储存储在磁盘文件中
Spillable Memory Channel	数据存储存储在内存中和磁盘上，当内存队满了，会持久化到磁盘文件（当前试验性的，不建议生产环境使用）

Memory Channel的特点是速度快，高吞吐，缺点是容量有限，Agent死掉之后数据丢失；FileChannel的特点是容量高，数据完整性强，即使Agent死掉，也可以恢复，缺点就是速度慢，单位时间吞吐量较低。JDBC Channel则可以提供持久化支持。Spillable Memory Channel由于是实验的，不可贸然用于生产环境。

而Memory，JDBC，File三种Channel的任何一种，又不可能同时满足大缓存、高吞吐的系统要求，必须结合公司实际情况，Channel进行个性化订制。当然，系统运行的初期在对处理速度和稳定性要求不高情况下可以先使用他们上述三种。

**四、预期效果**

随着大数据技术的发展，其对于海量数据快速分析和处理的优势进一步显现。可以预期，日志集系统的运行，将使得以往需要大量人力物力时间成本来完成的工作，可以在短时间内高效达成，至有以下几点优势：

1、异构系统海量日志的同时收集处理。

2、发现和预防问题的实时性大幅提升。

3、异常处理速度的大幅提升。

4、系统优化分析速度的大幅提升。

结语：大数据并非高不可攀，最大的拦路虎就是行动的意愿。

```
<script>
  (adsbygoogle = window.adsbygoogle || []).push({});
</script>
<p> <br> </p>
<p> 参考资料: </p>
<p> <a href="https://link.hacpai.com/forward?goto=http%3A%2F%2Fdongxicheng.org%2Fs
arch-engine%2Flog-systems%2F" target="_blank" rel="nofollow ugc">开源日志系统比较</a>
</p>
<p> <a href="https://link.hacpai.com/forward?goto=http%3A%2F%2Fmy.oschina.net%2Fleej
n2005%2Fblog%2F288136" target="_blank" rel="nofollow ugc">Flume NG 简介及配置实战</
> </p>
<p> <a href="https://link.hacpai.com/forward?goto=http%3A%2F%2Fwww.aboutyun.com%2
thread-7415-1-1.html" target="_blank" rel="nofollow ugc">Flume内置channel,source,sink三
件介绍</a> </p>
<p> <a href="https://link.hacpai.com/forward?goto=http%3A%2F%2Ftech.meituan.com%2F
t-log-system-arch.html" target="_blank" rel="nofollow ugc">基于Flume的美团日志收集系统(一
架构和设计</a> </p>
<p> <a href="https://link.hacpai.com/forward?goto=http%3A%2F%2Fblog.newitfarmer.com
2Fbig_data%2Fstreams%2Fstorm%2F15470%2Fpost-flume-ngkafkastormhdfs-%25E5%25A
%259E%25E6%2597%25B6%25E7%25B3%25BB%25E7%25BB%259F%25E6%2590%25AD%25
5%25BB%25BA" target="_blank" rel="nofollow ugc">flume-ng+Kafka+Storm+HDFS 实时系
搭建</a> </p>
```