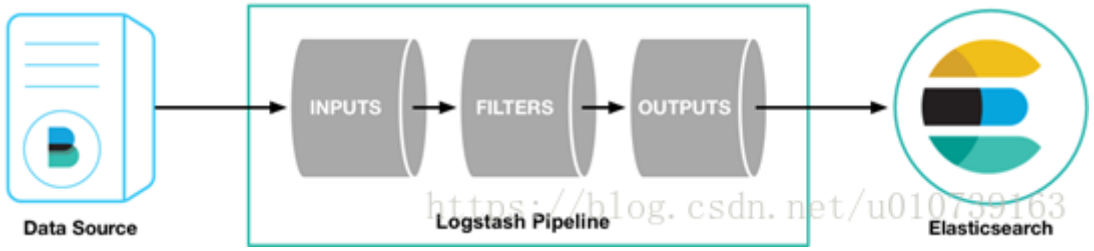
一、ELK介绍

对于日志来说，最常见的需求就是收集、存储、查询、展示，开源社区正好有相对应的开源项目：logstash（收集）、elasticsearch（存储+搜索）、kibana（展示），我们将这三个组合起来的技术称之为ELKStack，所以说ELKStack指的是Elasticsearch、Logstash、Kibana技术栈的结合。

二、Logstash简介

Logstash 是一款强大的数据处理工具，它可以实现数据传输，格式处理，格式化输出，还有强大的插件功能，常用于日志处理。

三、Logstash工作原理



输入，以下是常见得输入内容

1) file：从文件系统上的文件读取，与UNIX命令非常相似 tail -0F

2) syslog：在已知端口上侦听syslog消息进行解析

3) redis：使用redis通道和redis列表从redis服务器读取。Redis通常用作集中式Logstash安装中的“代理”，该安装将Logstash事件从远程Logstash“托运人”排队。

4) beats：处理 Beats发送的事件,beats包括filebeat、packetbeat、winlogbeat。

过滤，以下是常见得过滤器

1) grok：解析并构造任意文本。Grok是目前Logstash中将非结构化日志数据解析为结构化和可查询内容的最佳方式。Logstash内置了120种模式，您很可能会找到满足您需求的模式！

2) mutate：对事件字段执行常规转换。您可以重命名，删除，替换和修改事件中的字段。

3) drop：完全删除事件，例如调试事件。

4) clone：制作事件的副本，可能添加或删除字段。

5) geoip：添加有关IP地址的地理位置的信息（也在Kibana中显示惊人的图表！）

输出，以下是常见得输出内容

1) elasticsearch：将事件数据发送给Elasticsearch。如果您计划以高效，方便且易于查询的格式保存数据..... Elasticsearch是您的最佳选择
2) file：将事件数据写入磁盘上的文件。
3) graphite：将事件数据发送到graphite，这是一种用于存储和绘制指标的流行开源工具。http://graphite.readthedocs.io/en/latest/
4) statsd：将事件数据发送到statsd，这是一种“侦听统计信息，如计数器和定时器，通过UDP发送并将聚合发送到一个或多个可插入后端服务”的服务。如果您已经在使用statsd，这可能对您有用！

编解码器
编解码器基本上是流过滤器，可以作为输入或输出的一部分运行。使用编解码器可以轻松地将消息传输与序列化过程分开。流行的编解码器包括json, multiline等。
json：以JSON格式编码或解码数据。
multiline：将多行文本事件（例如java异常和堆栈跟踪消息）合并到一个事件中

四、Logstash优点

1、可伸缩性
节拍应该在一组Logstash节点之间进行负载平衡。
建议至少使用两个Logstash节点以实现高可用性。
每个Logstash节点只部署一个Beats输入是很常见的，但每个Logstash节点也可以部署多个Beats输入，以便为不同的数据源公开独立的端点。

2、弹性
Logstash持久队列提供跨节点故障的保护。对于Logstash中的磁盘级弹性，确保磁盘冗余非常重要。对于内部部署，建议您配置RAID。在云或容器化环境中运行时，建议您使用具有反映数据SLA的复制策略的永久磁盘。

3、可过滤
对事件字段执行常规转换。您可以重命名，删除，替换和修改事件中的字段。

4、可扩展插件生态系统，提供超过200个插件，以及创建和贡献自己的灵活性。

五、Logstash缺点

Logstash耗资源较大，运行占用CPU和内存高。另外没有消息队列缓存，存在数据丢失隐患。

Logstash使用详解

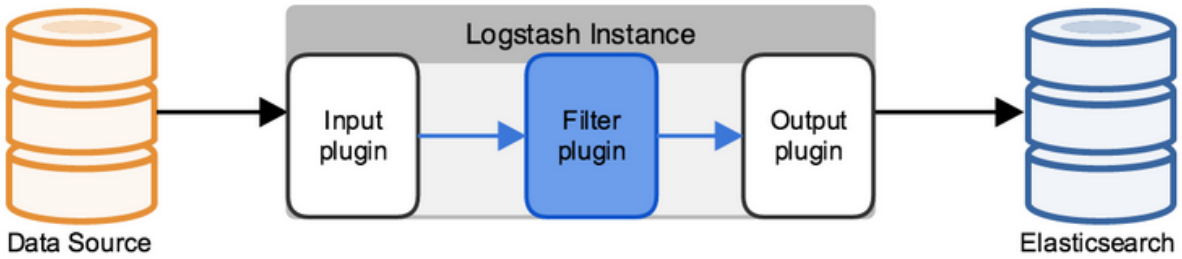
Logstash是一款轻量级的日志搜集处理框架，可以方便的把分散的、多样化的日志搜集起来，并进行自定义的处理，然后传输到指定的位置，比如某个服务器或者文件。

工作原理

Logstash使用**管道方式**进行日志的搜集处理和输出。有点类似*NIX系统的管道命令 **xxx | ccc | ddd**，xxx执行完了会执行ccc，然后执行ddd。

在logstash中，包括了三个阶段:

输入input --> 处理filter（不是必须的） --> 输出output



每个阶段都由很多的插件配合工作，比如file、elasticsearch、redis等等。

每个阶段也可以指定多种方式，比如输出既可以输出到elasticsearch中，也可以指定到stdout在控制台打印。

由于这种插件式的组织方式，使得logstash变得易于扩展和定制