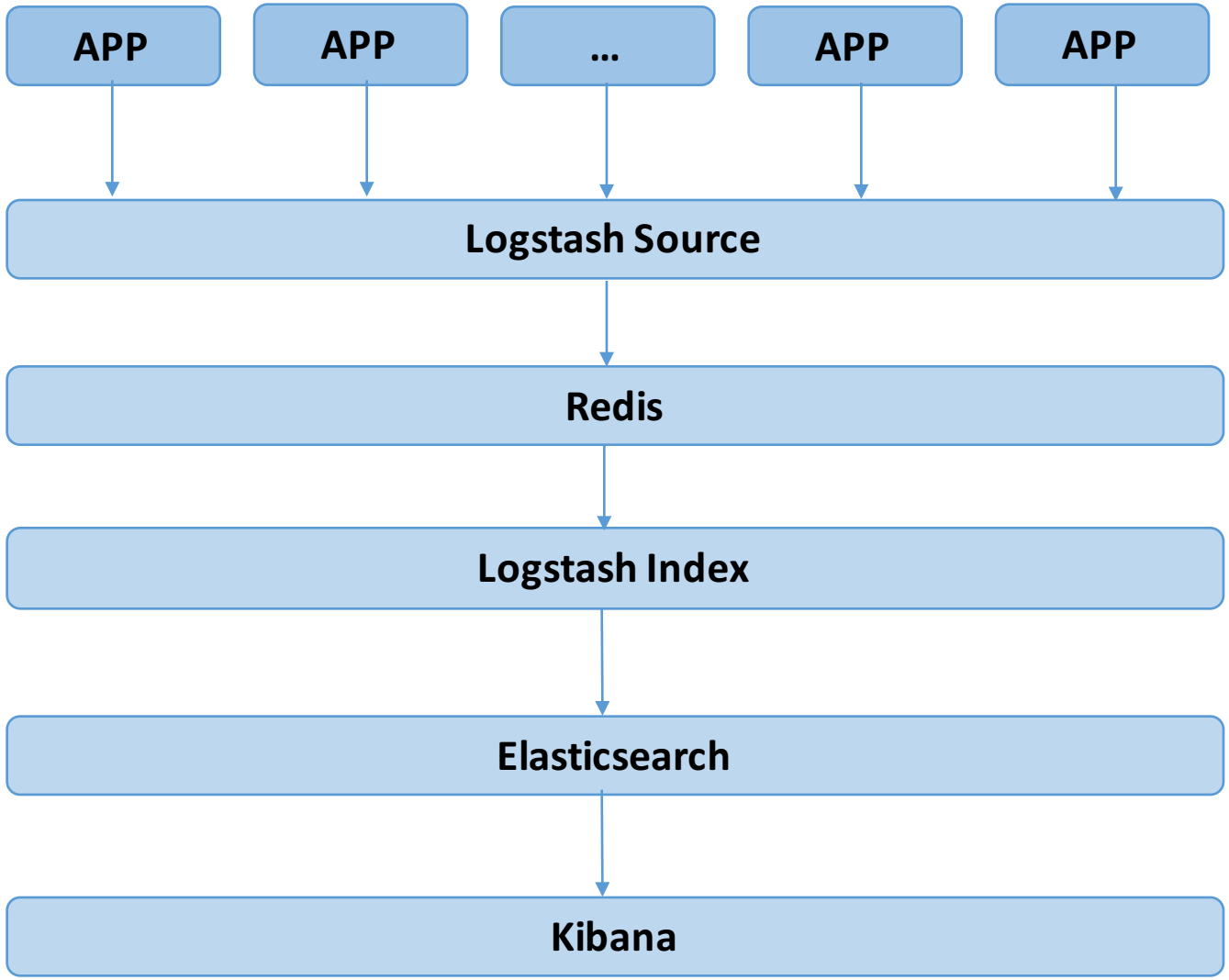
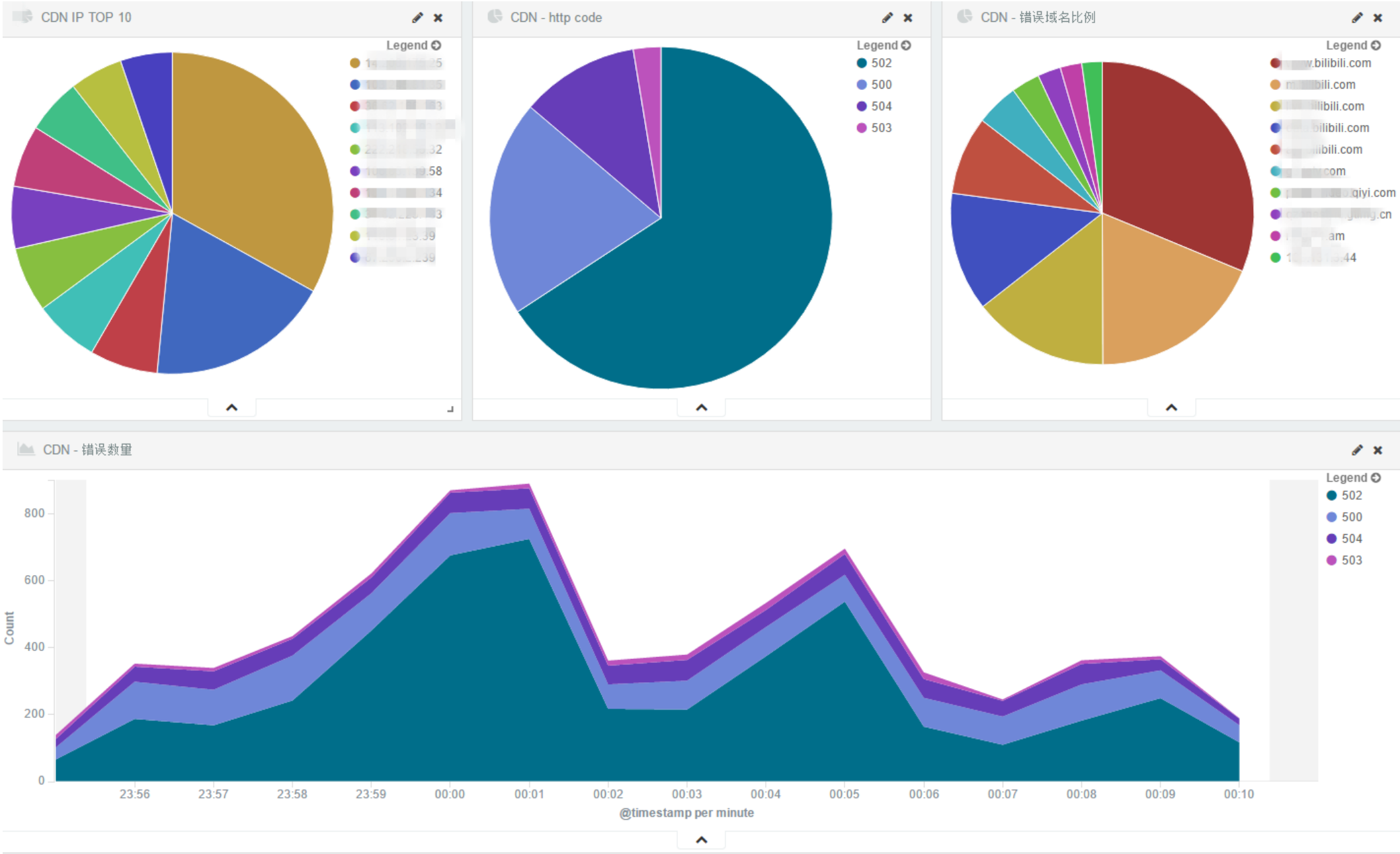
B站日志流实施之路

—宋金刚


BOOM





- 插件式组织方式，易于扩展和控制
- 数据源多样不仅限于日志文件，数据处理操作更丰富，可自定义（过滤，匹配过滤，转变，解析.....）
- 可同时监控多个数据源（input插件多样），同时也可将处理过的数据同时有不同多种输出（如stdout到控制台，同时存入elasticsearch）
- 4.安装简单，使用简单，结构也简单，所有操作全在配置文件设定，运行调用配置文件即可
- 管道式的dataSource——input plugin——filter plugin——output plugin——dataDestination
- 7.有一整套的EKL日志追踪技术栈，可收集处理（logstash），存储管理搜索（elasticsearch），图形显示分析（kibana）

- Logstash由Ruby开发当日志量大时 Logstash Source往往会成为整个日志流的瓶颈，为了避免瓶颈增加的ELK的复杂程度，运维成本更大。

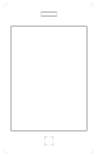
- 对于上下文检索实现困难，对于某些问题查找很多时候需要上下文检索功能，在linux终端，可以很快捷的使用tail head grep等命令实现，但是对于使用ELK做日志分析，显得乏力。

大数据日志搜集组件 Flume

高可用，分布式。

声明式配置，可动态更新配置。

海量日志收集、聚合、传输系统。



Flume的三层架构 Source Channel Sink

Source ：

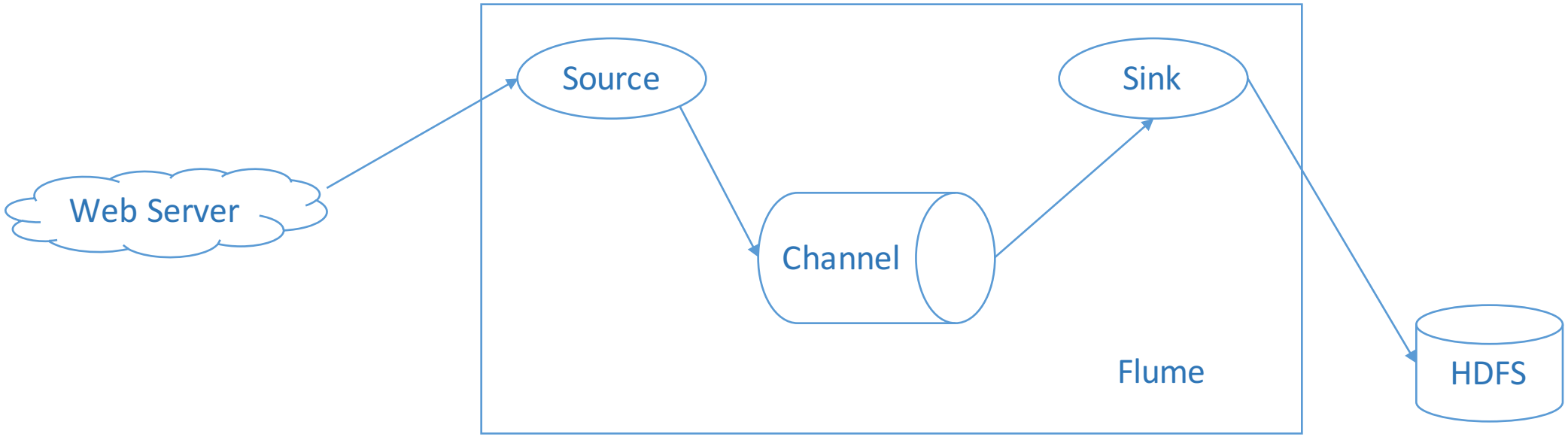
负责接收数据支持多种数据源输入 如：kafka、syslog、netcat、HTTP等
并且提供拦截器（Interceptors）对数据流做相应操作。
Source将搜集到的数据存放在Channel中。

Channel ：

相当于flume的队列系统，负责数据临时存储，Channel支持多种数据存储方式 如：Memory 、File、kafka、jdbc 等。

Sink ：

从Channel提取数据并将日志输出，支持HDFS、ElasticSearch、kafka等



Flume怎么样与ELK结合？

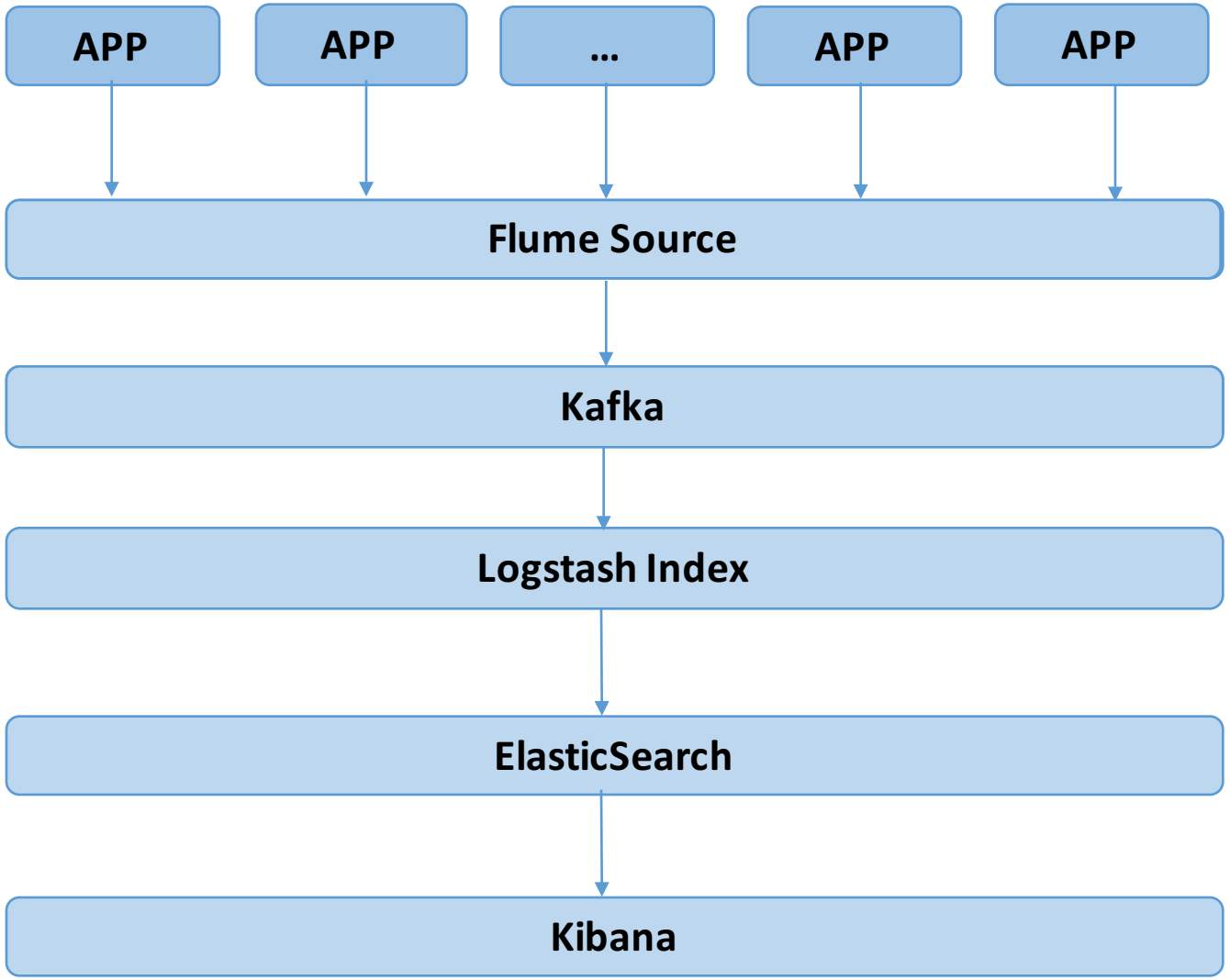
Flume日志传输的高效率

Logstash的强大日志处理功能。

Kafka的高并发能力。

ElasticSearch的检索功能

Kibana的展示功能



Flume 替换Logstash做日志搜集

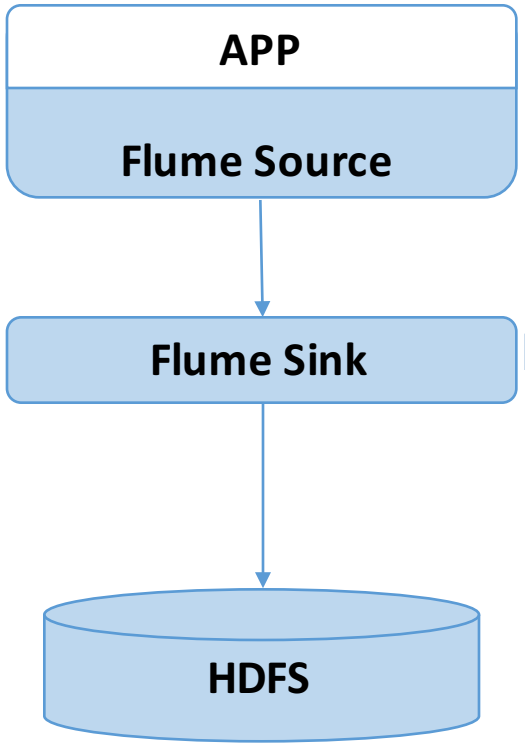
Kafka 替换redis做日志队列

Logstash Index做日志过滤并写入ES

这样就完了？

哦，对，还有大数据的日志流系统

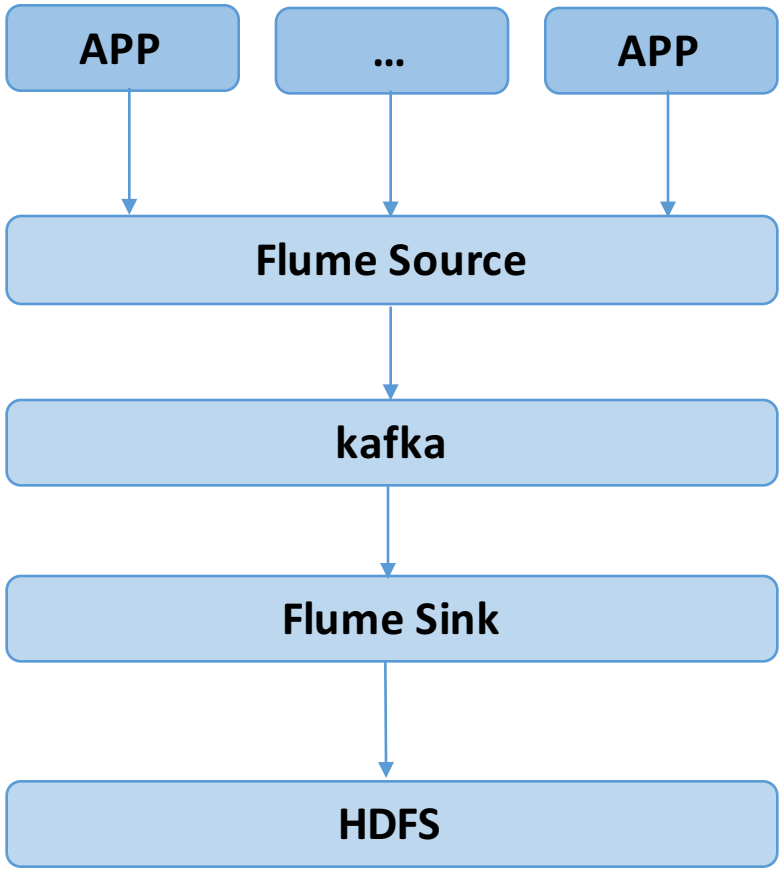
最初的大数据日志流系统



部署复杂，每台业务机器都需要有flume进程

Flume Channel队列堆积超过内存大小导致日志丢失

HDFS经常有大量的数据读写，flume写入会变得缓慢。



移除APP机器上的Flume并规定统一格式写入对应的Flume端口

Flume Source只需开放相应端口，将接收到的数据写入kafka

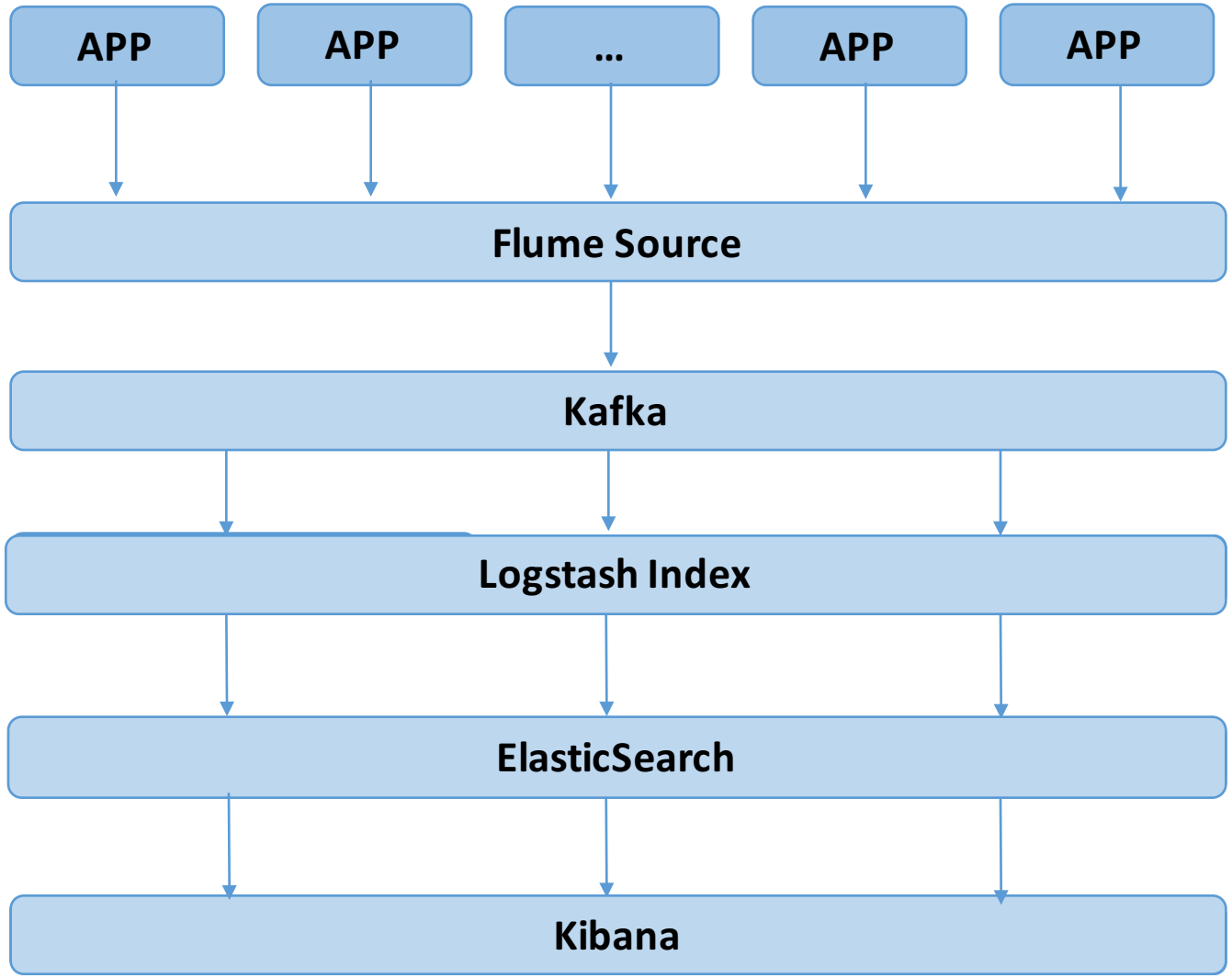
搭建kafka集群作为日志流队列

Sink端使用lzo格式压缩日志写入HDFS

再整合

两套日志流系统需要维护

ELK的日志同时要写入HDFS



THANKS

哔哩哔哩 - (゜- ゜)つロ 乾杯~ - bilibili