开讲啦 | EFB运行日志分析技术平台ELK介绍

2017-07-05 东航IT百分百

1. ELK介绍

1.1. 什么ELK

ELK是软件集合 Elasticsearch、Logstash、Kibana的简称，由这三个软件及其相关的组件可以打造大规模日志实时处理系统。

其中，Elasticsearch 是一个基于Lucene 的、支持全文索引的分布式存储和索引引擎，主要负责将日志索引并存储起来，方便业务方检索查询。

Logstash 是一个日志收集、过滤、转发的中间件，主要负责将各条业务线的各类日志统一收集、过滤后，转发给 Elasticsearch 进行下一步处理。

Kibana 是一个可视化工具，主要负责查询Elasticsearch 的数据并以可视化的方式展现给业务方，比如各类饼图、直方图、区域图等。

所谓“大规模”指的是 ELK 组成的系统以一种水平扩展的方式每天支持收集、过滤、索引和存储 TB 规模以上的各类日志（注：1TB = 1024GB ）。

---

2. ELK核心组件的介绍和应用

在ELK日志平台中能个熟悉ELasticSearch、Logstash 、Kibnan 三个组件能够帮我们更好去运用它。

2.1. ElasticSearch

ElasticSearch是一个基于Lucene的搜索服务器。它提供了一个分布式多用户能力的全文搜索引擎，基于RESTfulweb接口。Elasticsearch是用Java开发的，并作为Apache许可条款下的开放源码发布，是当前流行的企业级搜索引擎。设计用于云计算中，能够达到实时搜索，稳定，可靠，快速，安装使用方便。

我们建立一个网站或应用程序，并要添加搜索功能，但是想要完成搜索工作的创建是非常困难的。我们希望搜索解决方案要运行速度快，我们希望能有一个零配置和一个完全免费的搜索模式，我们希望能够简单地使用JSON通过HTTP来索引数据，我们希望我们的搜索服务器始终可用，我们希望能够从一台开始并扩展到数百台，我们要实时搜索，我们要简单的多用户，我们希

望建立一个云的解决方案。因此我们利用Elasticsearch来解决所有这些问题以及可能出现的更多其它问题。

2.2. Logstash

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。

当然它可以单独出现，作为日志收集软件，你可以收集日志到多种存储系统或临时中转系统，如MySQL，redis，kakfa，HDFS, lucene，solr等并不一定是ElasticSearch。

Logstash使用管道方式进行日志的搜集处理和输出。在logstash中，包括了三个阶段: 输入input –> 处理filter（不是必须的） –> 输出output

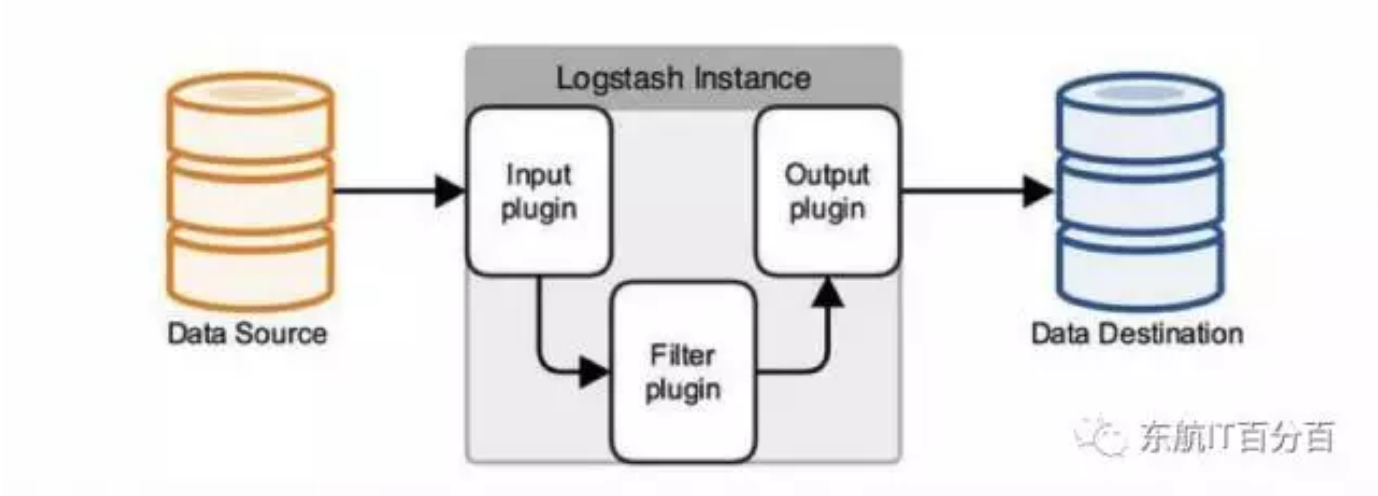

图1 logstash的三个阶段

2.3. Kibana

Kibana是一个开源的分析与可视化平台，设计出来用于和Elasticsearch一起使用的。你可以用kibana搜索、查看、交互存放在Elasticsearch索引里的数据，使用各种不同的图表、表格、地图等kibana能够很轻易地展示高级数据分析与可视化。

Kibana让我们理解大量数据变得很容易。它简单、基于浏览器的接口使你能快速创建和分享实时展现Elasticsearch查询变化的动态仪表盘。安装Kibana非常快，你可以在几分钟之内安装和开始探索你的Elasticsearch索引数据——-——-不需要写任何代码，没有其他基础软件依赖

2.3.1. Discover

您可以从 Discover（发现）页面以交互的方式来探索数据。您可以访问每个索引中与所选索引模式匹配的每个文档。您可以提交搜索查询，过滤搜索结果以及查看文档数据。您还可以查看与搜

索查询匹配的文档数，并获取字段值统计信息。如果为所选索引模式配置了时间字段，文档随时间的分布被显示在页面顶部的直方图中。

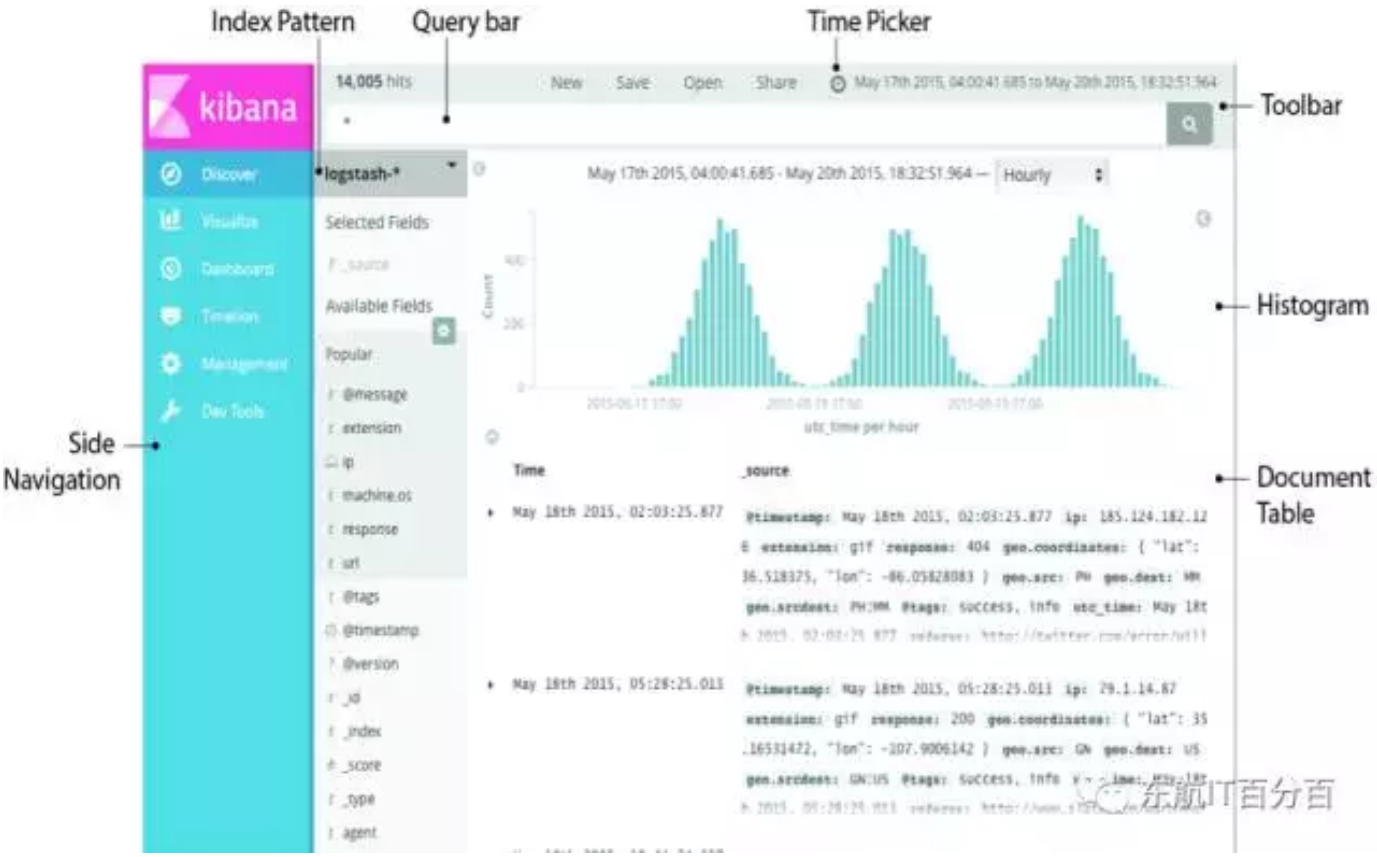

图2 Discover（发现）页面

2.3.2. Visible

Visualize（可视化）可以让你在 Elasticsearch索引中的数据上创建可视化。您也可以构建 Dashboard（仪表盘）来展示相关的可视化。

Kibana 的可视化是基于elsticsearch 的查询基础之上的，通过运用一系列的 elasticsearchaggregations（聚合）来提取以及处理数据，你也可以用创建图表的方式来展示你所需要的数据趋势、峰值、低值。

您可以从 Discover（发现）中已保存的Search（搜索）创建可视化，或者从新的搜索查询开始。

Area Charts（面积图）	可视化几个不同组的总贡献。
Data Table（数据表）	显示组合聚合的原始数据。
Line Charts（折线图）	比较不同的组。
Markdown Widget（小部件）	显示自由格式信息或说明。
Metric（度量）	显示单个数字。
Pie Charts（饼图）	显示每个资源相对全部的贡献。
Pie Charts（饼图）	将单词显示为云，其中字的大小对应于其重要性。
Tile Maps（平铺地图）	将聚合的结果与地理位置相关联。
Timeseries	计算和组合来自多个时间序列数据集的数据。
Vertical Bar Charts（垂直条形图）	在条形图中绘制图形值。

图3 创建可视化

2.3.3. DashBoard

仪表盘:一个 Kibanadashbooard （Kibana 仪表盘）能让你自由排列一组已保存的可视化。然后你可以保存这个仪表板，用来分享或者重载。

2.3.4. Timelion

Timelion :是一个时间序列数据可视化工具，使您能够在一个可视化中组合完全独立的数据源。它由一个简单的表达式语言驱动，用于检索时间

序列数据，执行计算来挑选复杂问题的答案，并可视化结果。

例如:Timelion 使您能够轻松获得以下问题的答案：

每个唯一的用户在一段时间内查看了多少次页面？

本周五和上周五之间的流量有什么区别？

日本有多少百分比的人口今天来到我的网站？

标准普尔 500 指数的 10 天移动均线是多少？

过去 2 年内收到的所有搜索请求的累积和是多少？

要开始构建时间序列可视化，请单击侧面导航中的 Timelion 并运行教程。 Timelion 表达式语言的文档是内置的。

2.3.5. DevTools

Console plugin（控制台插件）提供了一个 UI 来与 Elasticsearch 的 REST API 进行交互。 控制台有两个主要方面 : editor（编辑器），编写

对 Elasticsearch 的请求以及response（响应）窗格的地方，并且显示对请求的响应。

2.3.6. Management

管理应用程序是您执行 Kibana 的运行时配置的位置，包括索引模式的初始设置和持续配置，调整 Kibana 本身行为的高级设置，以及您可以在整个 Kibana 中保存的各种 "对象"，例如搜索（searches），可视化（visualizations）和仪表盘（dashboards）。

---

3. ELK日志分析平台示例

同步数据库中DispLog数据

Kibana的默认访问端口是:5601

本地访问:localhost:5601

3.1. 界面

图4 界面

3.2. 做出的网络饼状图:

图5 网络饼状图

---

4. ELK常用架构

随着ELK日志平台的发展,根据业务需求出现很多不同的架构,可以跟据企业业务需求做出相应的选择

4.1. 架构一

如图6，这是最简单的一种ELK架构方式。优点是搭建简单，易于上手。缺点是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。建议供学习者和小规模集群使用。

此架构首先由Logstash分布于各个节点上搜集相关日志、数据，并经过分析、过滤后发送给远端服务器上的Elasticsearch进行存储。 Elasticsearch将数据以分片的形式压缩存储并提供多种API供用户查询，操作。用户亦可以更直观的通过配置Kibana Web Portal方便的对日志查询，并根据数据生成报表

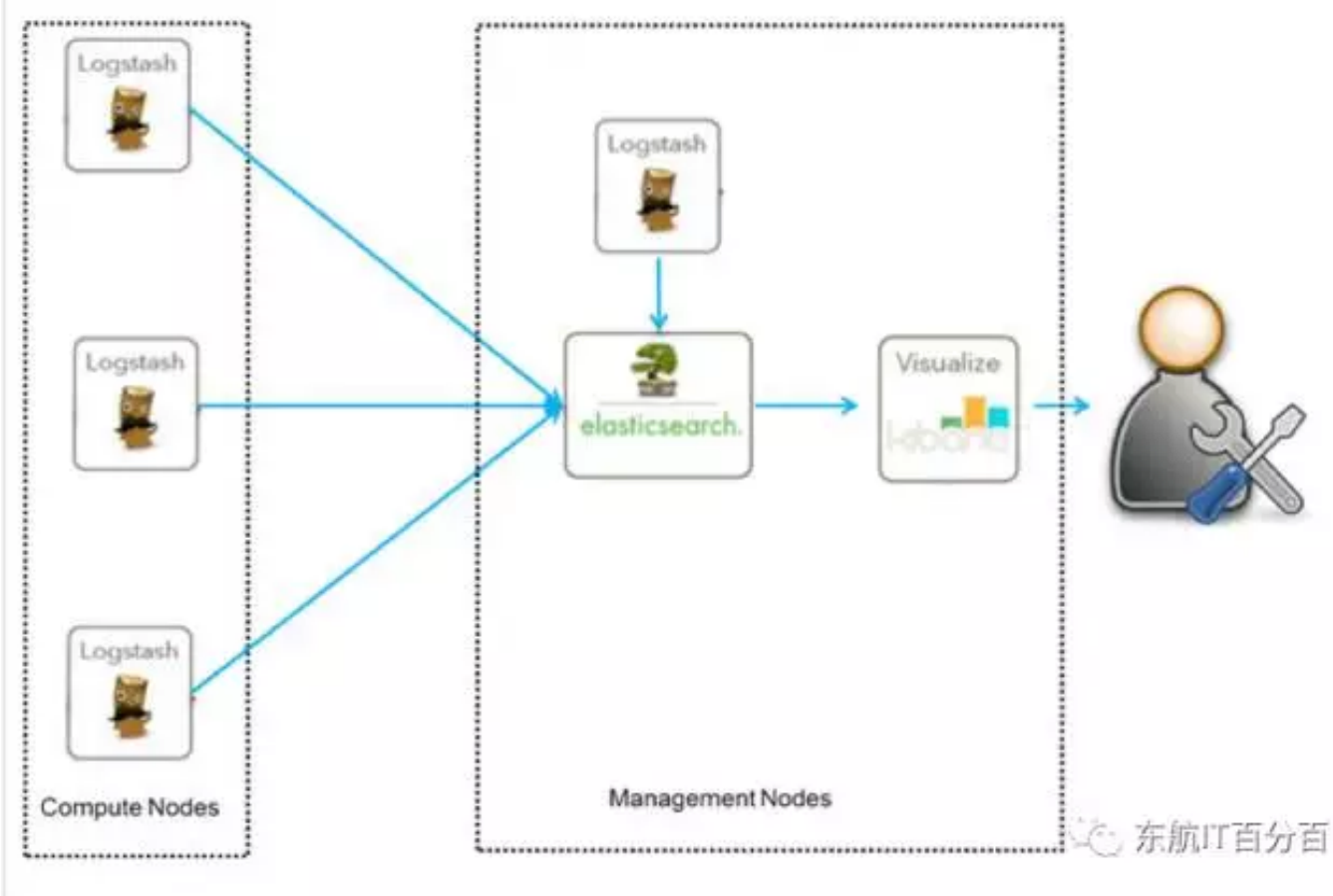

图6 ELK架构一

4.2. 架构二

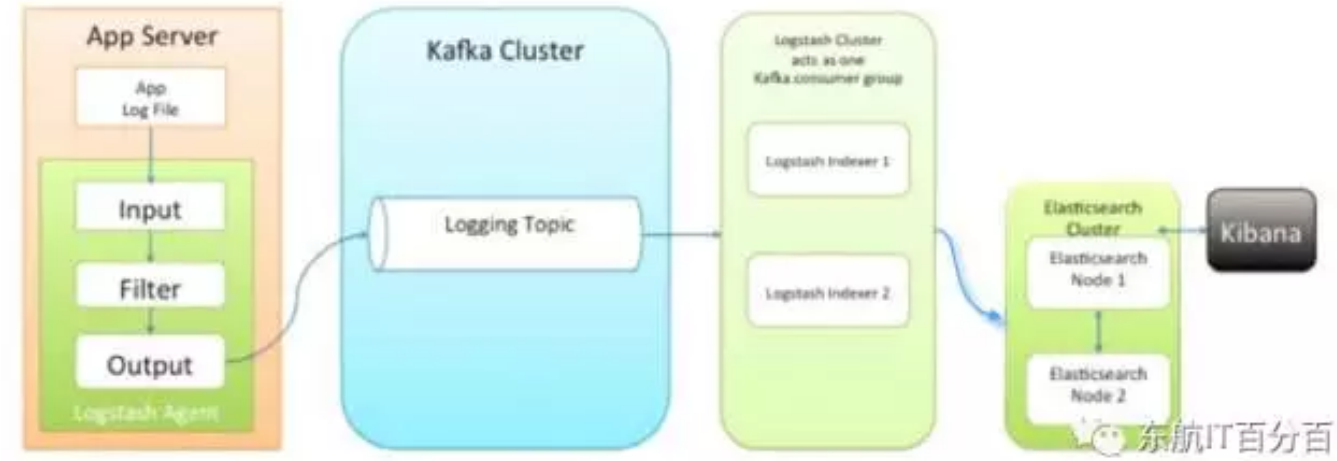

图7 ELK架构二

如图7 引入了消息队列机制，位于各个节点上的Logstash Agent先将数据/日志传递给Kafka（或者Redis），并将队列中消息或数据间接传递给Logstash，Logstash过滤、分析后将数据传 递给Elasticsearch存储。最后由Kibana将日志和数据呈现给用户。因为引入了Kafka（或者Redis）,所以即使远端 Logstash server因故障停止运行，数据将会先被存储下来，从而避免数据丢失。

4.3. 架构三

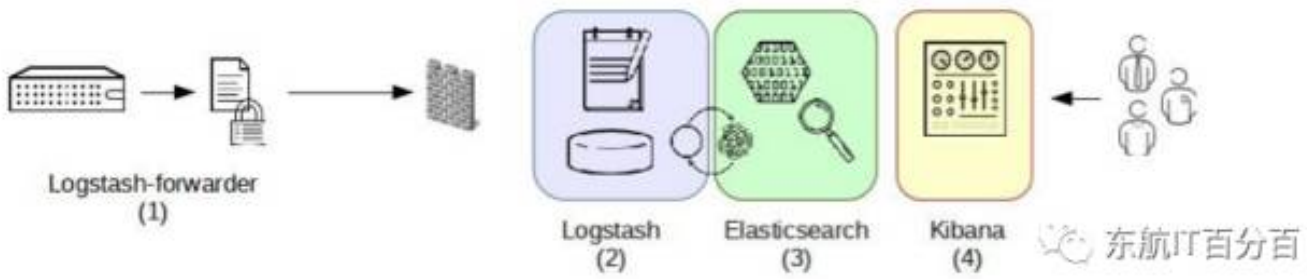

图8 ELK架构三

引入了Logstash-forwarder。首先，Logstash-forwarder将日志数据搜集并统一发送给主节点上的Logstash，Logstash分析、过滤日志数据后发送至Elasticsearch存储，并由Kibana最终将数据呈现给用户。

这种架构解决了Logstash在各计算机点上占用系统资源较高的问题。经测试得出，相比Logstash，Logstash-forwarder 所占用系统CPU和MEM几乎可以忽略不计。另外，Logstash-forwarder和Logstash间的通信是通过SSL加密传输，起到了安全保 障。如果是较大集群，用户亦可以如结构三那样配置logstash集群和Elasticsearch集群，引入High Available机制，提高数据传输和存储安全。更主要的配置多个Elasticsearch服务，有助于搜索和数据存储效率。但在此种架构下发现 Logstash-forwarder和Logstash间通信必须由SSL加密传输，这样便有了一定的限制性。

4.4. 架构四

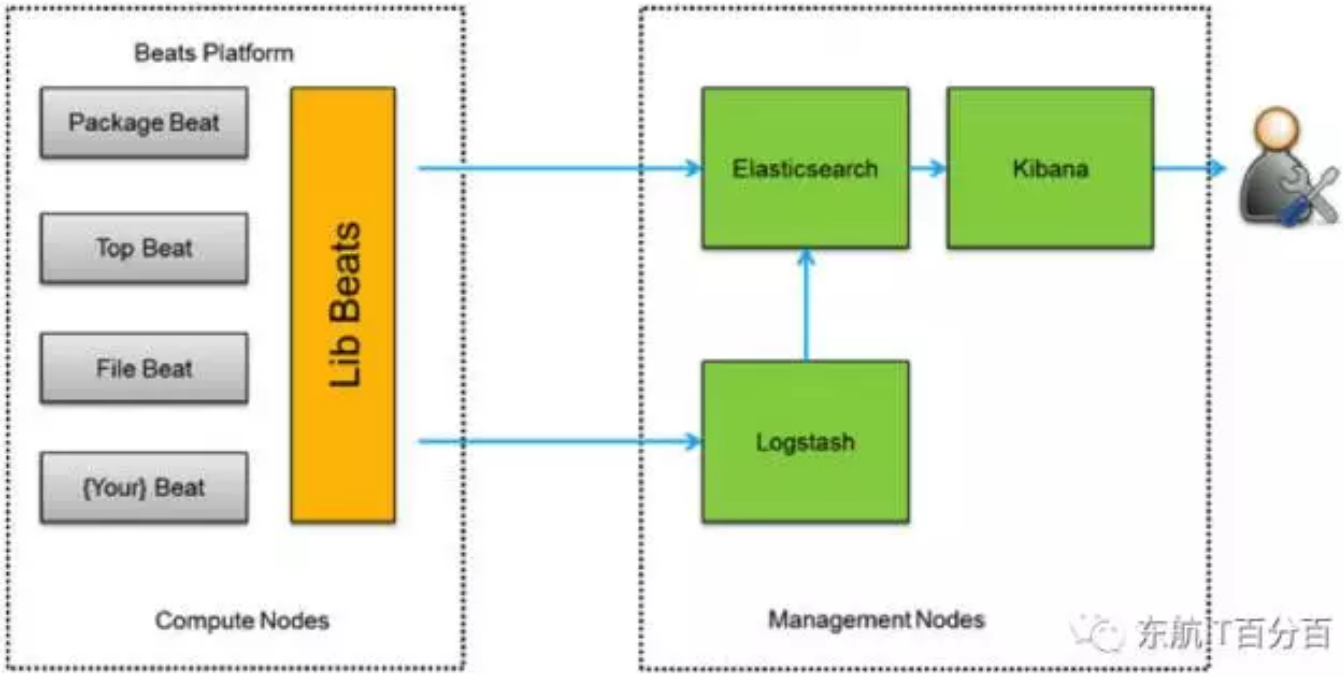

图9 ELK架构四

图9 将Logstash-forwarder替换为Beats。经测试，Beats满负荷状态所耗系统资源和Logstash-forwarder相当，但其扩展性和灵活性有很大提高。Beatsplatform目前包含有Packagebeat、Topbeat和Filebeat三个产品，均为Apache 2.0 License。同时用户可根据需要进行二次开发。

这种架构原理基于第三种架构，但是更灵活，扩展性更强。同时可配置Logstash 和Elasticsearch集群用于支持大集群系统的运维日志数据监控和查询。

- END -

信息部运行产品部 胡超华（审） 胡博文（文）