ELK由ElasticSearch，Logstash和Kiabana三个开源工具组成。

一，ELK概述

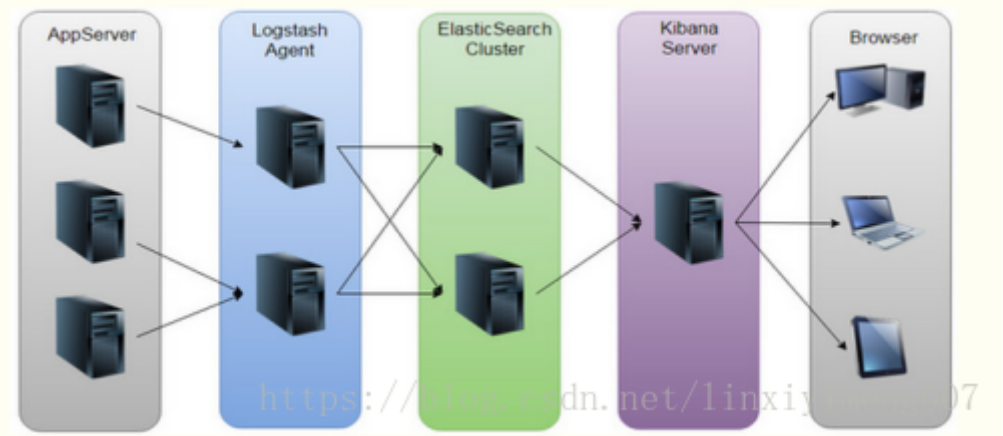
1 ，ELK 简介

ElasticSearch：是个开源分布式搜索引擎，它的特点有：分布式，零配置，自动发现，索引自动分片，索引副本机制， restful风格接口，多数据源，自动搜索负载等。

Logstash： 是一个完全开源的工具，他可以对你的日志进行收集，分析，并将其存储供以后使用。

Kibana：也是一个开源和免费的工具， Kibana可以为Logstash 和 ElasticSearch提供的日志分析友好的Web界面，可以帮助汇总，分析和搜索重要数据日志。

2 ，ELK 协议栈及体系结构



ELK其实并不是一款软件，而是一整套解决方案，是三个软件产品的首字母缩写，Elasticsearch，Logstash和Kibana。这三款软件都是开源软件，通常是配合使用，而且又先后归于弹性.co公司名下，故被简称为ELK协议栈

在需要收集日志的所有服务上部署logstash，Logstash收集应用服务器产生的日志，将日志收集在一起交给全文搜索服务ElasticSearch，而Kibana则从ES集群中查询数据生成图表，再返回给客户端浏览器。

二，ElasticSearch

2.1 ElasticSearch概述

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RESTful web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，用于设计云计算中，能够达到实时搜索，稳定，可靠，快速，安装使用方便。

Elasticsearch是一个实时分布式搜索和分析引擎与数据库对比来说：。现在大部分数据库在提取可用知识方面显得异常无能的确，它们能够通过时间戳或者精确匹配做过滤查询，但是它们不能够进行全文搜索，不能处理同义词，以及根据相关性给文档打分。而Elasticsearch能根据同一份数据生成分析和聚合的结果，最重要的是，它们在没有大量工作进程（线程）的情况下能做到对数据的实时处理。

2.2 ElasticSearch应用场景

互联网的时代，建立一个网站或应用程序，要添加搜索功能，实现站内搜索和站外搜索，但是想要完成搜索工作并保证运行的功能和性能是非常困难的。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应用，我们要实时搜索，我们要简单的多租户，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。

2.3 ElasticSearch基础概念

Elasticsearch有几个核心概念。从一开始理解这些概念会对整个学习过程有莫大的帮助。

接近实时（ NRT ） Elasticsearch是一个接近实时的搜索平台。这意味着，从索引一个文档直到这个文档能够被搜索到有一个轻微的延迟（通常是1秒）。

集群（ cluster ） 一个集群就是由一个或多个节点组织在一起，它们共同持有你整个的数据，并一起提供索引和搜索功能。一个集群由一个唯一的名字标识，这个名字默认就是“elasticsearch”。这个名字是重要的，因为一个节点只能通过指定某个集群的名字，

来加入这个集群。在产品环境中显式地设定，但是使用默认值来进行测试/开发也是不错的。

节点（节点） 一个节点是你集群中的一个服务器，作为集群的一部分，它存储你的数据，参与集群的索引和搜索功能。和集群类似，一个节点也是由一个名字标识的，默认情况下，这个名字是一个随机的漫威漫画角色的名字，这个名字会在启 的时候赋予节点。这个名字对于管理工作来说挺重要，因为在这个管理过程中，你会去确定网络中的哪些Elasticsearch集群中的哪些节点。 一个节点可以通过配置集群名称的来加入一个指定的集群。默认情况下，每个节点都会被安排加入到一个“elasticsearch”的集群中，这意味着，如果你在你的网络中启动了，并假定它们能够相互发现彼此，它们将会自动地形成并加入到“elasticsearch”的集群中。

在一个集群里，只要你想，可以拥有任意多个节点。而且，如果当前你的网络中没有运行任何Elasticsearch节点，这时启动一个节点，会默认创建并加入一个叫做“elasticsearch”的集群。

索引（索引）一个索引就是一个拥有几分相似特征。比如说，你可以有一个客户数据的索引，另一个产品目录的索引，还有一个订单数据的索引。一个索引由一个名字来标识（必须全部是小写字母的），并且当我们要对对应于这个，搜索，更新和删除的时候，都要使用到这个名字。在一个集群中，如果你想，可以定义任意多的索引。

类型（ type）在一个索引中，你可以定义一种或多种类型。一个类型是你的索引的一个逻辑上的分类/分区，其语义完全由你来定。 ，会为具有一组共同字段的。比如说，假设我们运你一个博客平台个人文库并且将你所有的数据存储到一个索引中。在这个索引中，你可以为用户数据定义一个类型，为博客数据定义另一个类型，当然，也可以为评论数据定义另一个类型。

文档 文件）一个文档是一个可被索引的。比如，你可以拥有某一个客户的文档，某一个产品的一个文档，当然，也可以拥有某个订单的一个文档。文档以JSON（Java脚本对象符号）格式来表示，而JSON是一个到处存在的互联网数据交互。在一个索引/类型里面，只要你想，你可以存储任意多的文档。注意，尽管一个文档，物理上存在于一个索引之中，文档必须被索引/赋予一个索引的类型。

分片和复制（碎片＆副本） 一个索引可以存储超出单个结点硬件限制的大量数据比如，一个具有。10亿文档的索引占据1TB的磁盘空间，而任一节点都没有这样大的;或者单个节点

处理搜索请求，响应太慢。 为了解决这个问题，Elasticsearch提供了将索引划分成多份，这些份就叫做分片。当你创建一个索引的时候，你可以指定你想要的分片。每个分片本身也是一个功能完善“索引”，这个“索引”可以被放置到集群中的任何。

分片之所以重要，主要有两方面的原因：

- 允许你水平分割/扩展你的内容容量
- 允许你在分片（潜在地，位于多个节点上）之上进行分布式的，并行的操作，进而提高性能/有效产出

至于一个分片怎样分布，它的文档怎样聚合回搜索请求，是完全由Elasticsearch管理的，对于作为用户的你来说，这都是透明的。
默认一个索引有 5个分片，每个分片都1个副本。

三，Kibana

3.1 Kibana概述

Kibana是一个开源的分析和可视化平台，旨在与Elasticsearch合作.Kibana提供搜索，查看和存储在Elasticsearch索引中的数据进行交互的功能。开发者或运维人员可以轻松地执行高级数据分析，并在各种图表，表格和地图中可视化数据。

四，Logstash

4.1 logstash概述

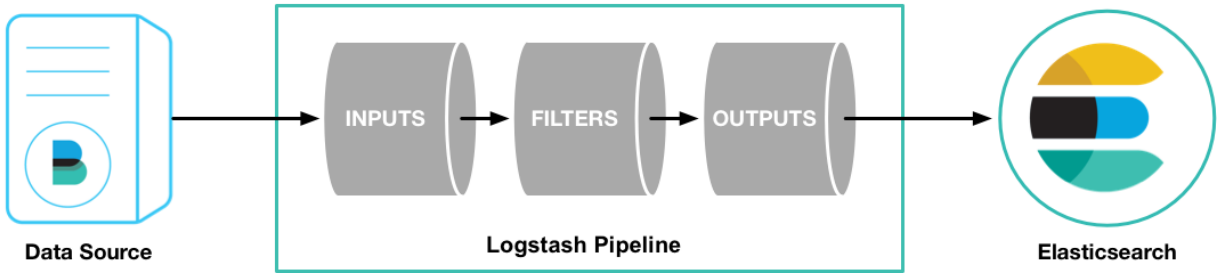
Logstash是一个开源的数据收集引擎，它具有备实时数据传输能力。它可以统一过滤来自不同源的数据，并按照开发者的制定的规范输出到目的地。

顾名思义， Logstash收集数据对象就是日志文件。由于日志文件来源多（如：系统日志，服务器日志等），且内容杂乱，不便于人类进行观察。因此，我们可以使用Logstash对日志文件进行收集和统一过滤，变成可读性高的内容，方便开发者或运维人员观察，从而有效的分析系统/项目运行的性能，做好监控和预警的准备工作等。

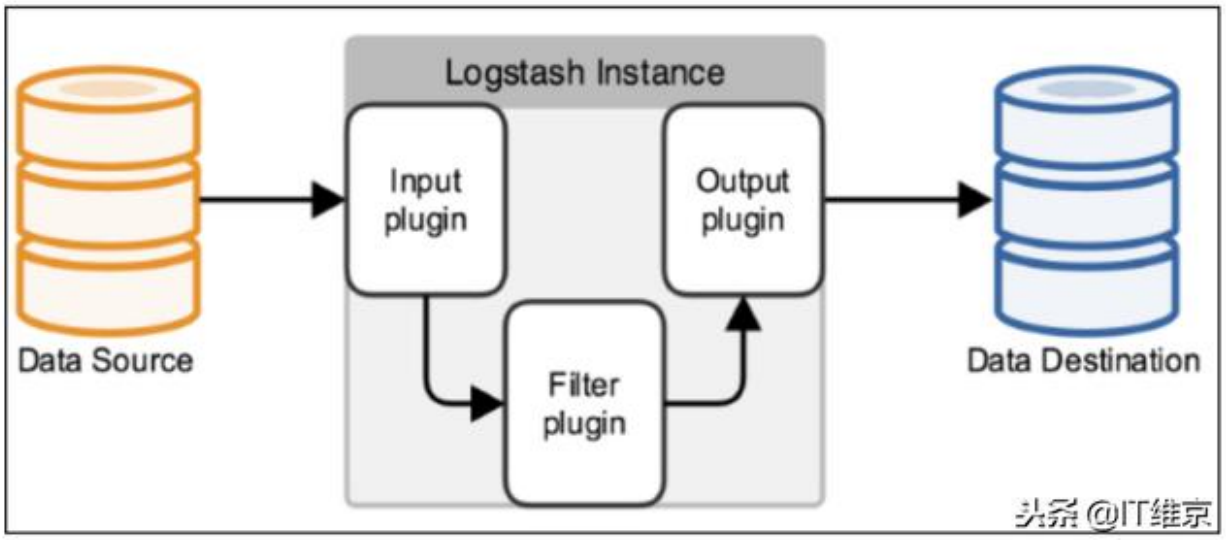
4.2 logstash组成结构

Logstash通过管道进行运作，管道有两个必要的元素，输入（输入）和输出（输出），还有一个可选的元素 - 过滤器（filter）。输入插件从数据源获取数据，过滤器插件根据。用户指定的数据格式修改数据，输出插件则将数据写入到目的地如下图：

Logstash的工作原理：



Logstash事件处理程序有三个阶段：输入→过滤器→输出。它是用于接收，处理和转发日志的工具。支持系统日志，Web服务器日志，错误日志，应用程序日志以及可以引发的所有日志类型。输入：将数据输入logstash。



一些常见的输入是：

文件：从文件系统文件中读取，类似于tial -f命令
Syslog：侦听端口514上的系统日志消息，并根据RFC3164标准进行解析
Redis：从redis服务中读取

节拍：从filebeat读取
过滤器：对数据进行操作的数据中间处理。

一些常用的过滤器是：

Grok：解析任意文本数据，Grok是Logstash最重要的插件。其主要功能是将文本格式的字符串转换为具体的结构化数据，以便与正则表达式一起使用。内置了120多种解析语法。

官方gro表达式：https：//github.com/logstash-plugins/logstash-patterns-core/tree/master/patterns

Grok在线调试：https：//grokdebug.herokuapp.com/

变异：转换字段。例如，删除字段、替换、修改、重命名，依此类推。

删除：丢弃某些事件但不处理它们。

克隆：复制事件，此字段还可以添加或删除字段。

Geoip：添加地理信息（用于前端kibana图形显示）

输出：输出是logstash处理管道的最终组件。事件可以在处理期间传递多个输出，但是一旦所有输出完成执行，事件将完成生命周期。
一些常见的编解码器：

Json：使用json格式对数据进行编码/解码。
多行：将来自多个事件的数据合并为一行。例如：java异常信息和堆栈信息。

官方资料：

Elasticsearch：
Https://www.elastic.co/cn/products/elasticsearch
Https://www.elastic.co/guide/en/elasticsearch/reference/5.6/index.html

Logstash：

Https://www.elastic.co/cn/products/logstash

Https://www.elastic.co/guide/en/logstash/5.6/index.html

Kibana：

Https://www.elastic.co/cn/products/kibana

Https://www.elastic.co/guide/en/kibana/5.5/index.html

Filebeat：

Https://www.elastic.co/cn/products/beats/filebeat

Https://www.elastic.co/guide/en/beats/filebeat/5.6/index.html

Elasticsearch 中文社区：

Https://elasticsearch.cn/