高可用 ELK 架构

李海滨 @ 新致软件

2018.07.20

什么是 ELK？

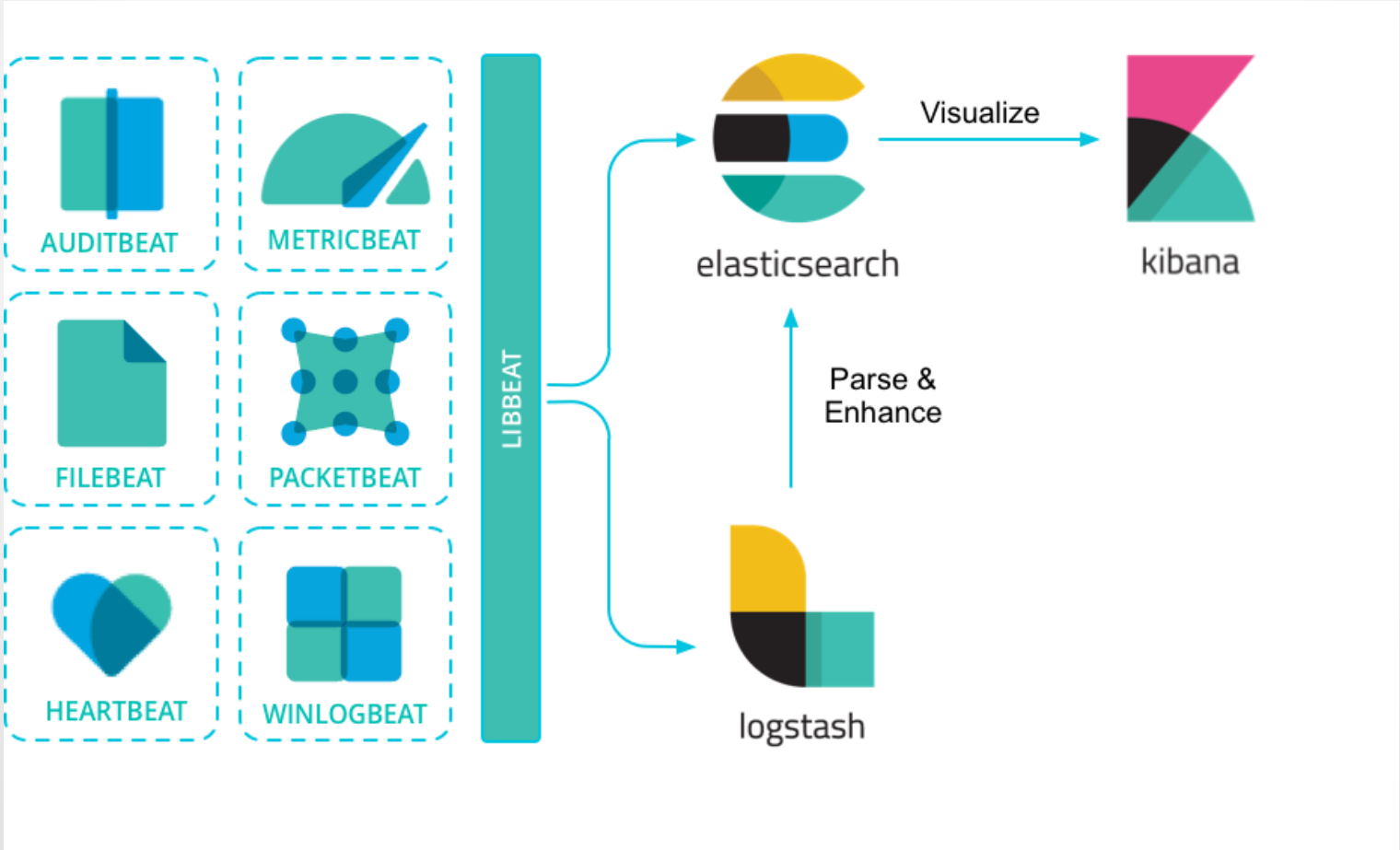
- Elasticsearch，高弹性开源全文检索及分析引擎，用于日志的存储、检索和统计分析。后文简称“ ES”。
- Logstash，开源数据收集引擎，可对日志数据进行再处理。
- Kibana，开源分析及可视化平台，只与 ES 搭配使用。

什么是 Beats？

- Filebeat
  elastic 官方的 Linux 文本日志采集客户端，可对日志做前期过滤、整合，发往指定的上游服务端。
- Winlogbeat
  elastic 官方的 Windows 日志采集客户端，可对日志做前期过滤、整合，发往指定的上游服务端。
- Metricbeat
  elastic 官方的资源采集客户端，可采集系统资源状态，以及常见的应用服务资源状态，例如： MySQL、 Redis， RabbitMQ 等。
- Packetbeat
  elastic 官方的网络流量采集客户端，可采集网络的
  ICMP、 TCP、 DNS、 HTTP， MySQL 等协议的数据包。

Beats 的 output

采集的数据可直接输出到 ES，或经过 logstash 二次处理后再写入 ES。
还支持往 Kafka、 Redis、 File 和 Console 输出。
但一个 beat agent 同时只能使用一种 output 方式。



Logstash 的 input 和 output

Logstash 可同时输入和输出多种数据源。

可获取的 input 源	可输出的 output 端
beats	csv
elasticsearch	elasticsearch
exec	email
file	exec
github	file
log4j	graphite
rabbitmq	http
redis	influxdb
tcp	kafka
syslog	mongodb
kafka	rabbitmq
websocket	zabbix
更多 https://www.elastic.co/guide/en/logstash/current/input-plugins.html	更多 https://www.elastic.co/guide/en/logstash/current/output-plugins.html

新致高可用 ELK 集群特性

- 服务自发现自注册，弹性伸缩。
- 服务状态自动监控，服务自愈。
- 系统性能监控，告警信息可发往邮箱、钉钉（DingTalk）、 zabbix 等。
- 热数据定时迁移到冷数据节点。
- 支持跨机房多集群分布式部署，只需 1 个 Kibana 可查看所有集群的日志数据。

涉及的应用服务

Consul （https://www.consul.io/）

Service Discovery and Configuration

Redis + Sentinel （https://redis.io/）

In-memory data structure store

Monit （https://mmonit.com/monit/）

System monitoring and error recovery.

Chrony （https://chrony.tuxfamily.org/）

A versatile implementation of the Network Time Protocol

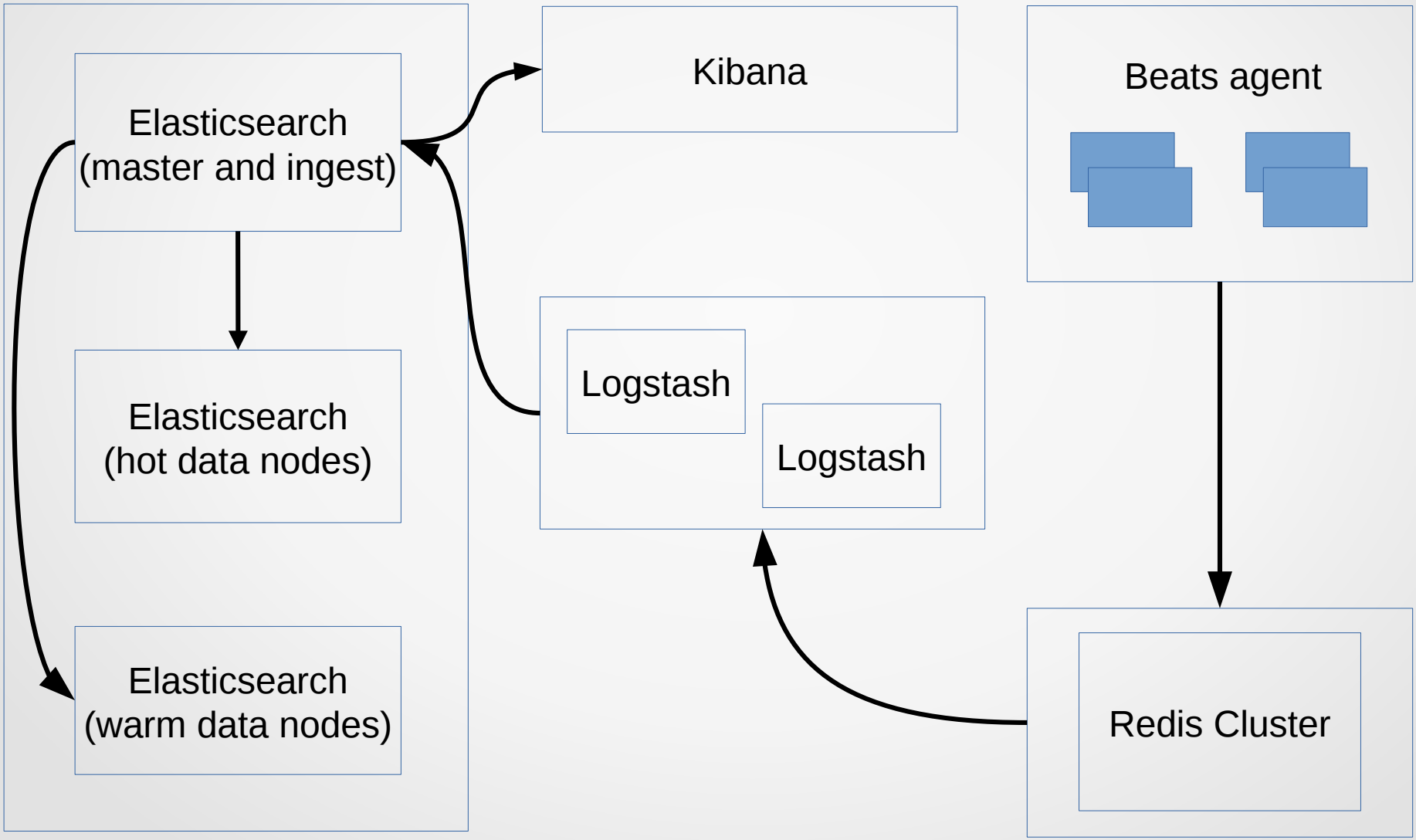
ELK Stack （https://www.elastic.co）

Reliably and securely take data from any source, in any format, and

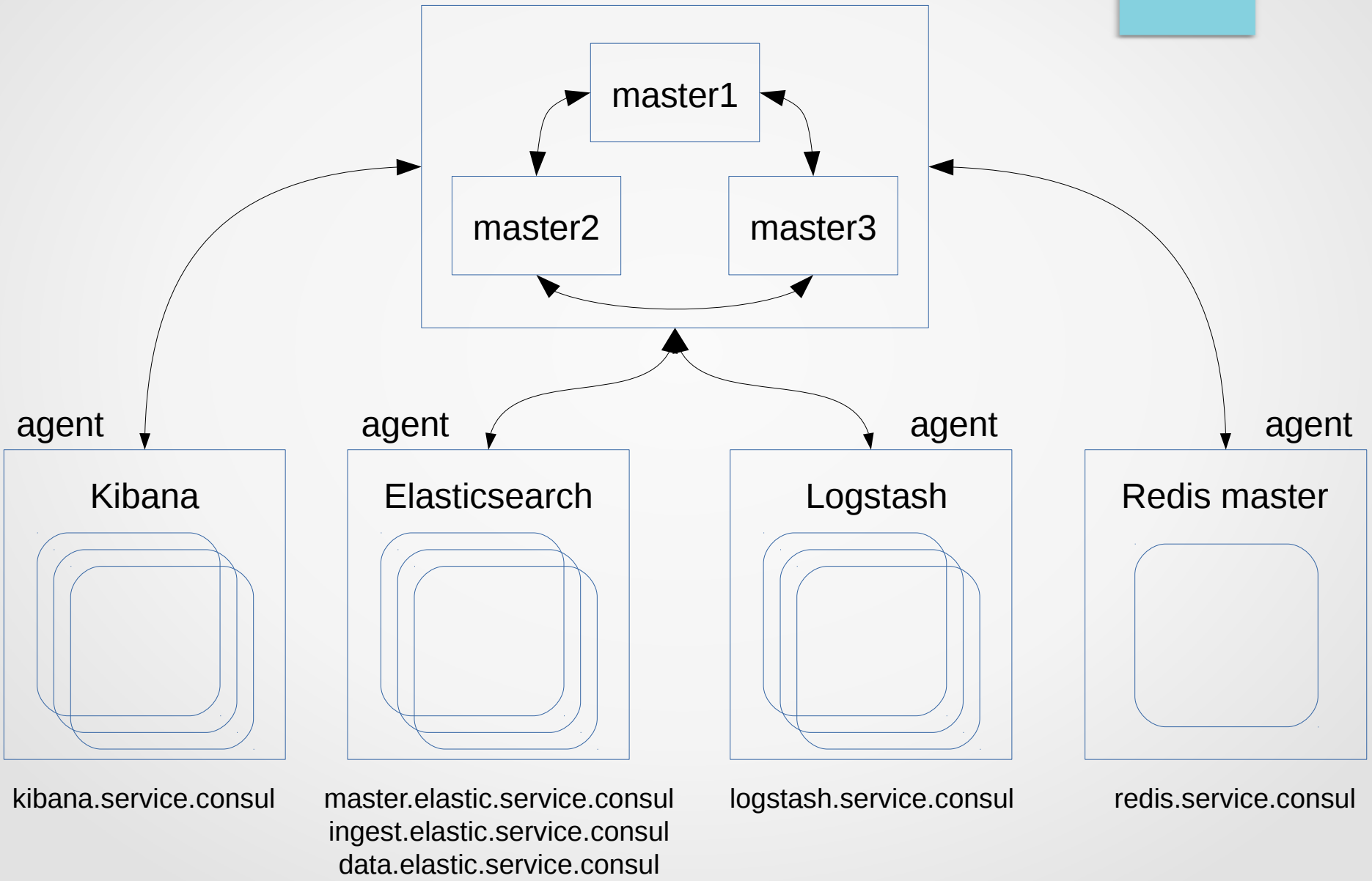
search, analyze, and visualize it in real time.

Beats （https://www.elastic.co/cn/products/beats）

ELK 流程图


Elasticsearch
(master and ingest)
Elasticsearch
(hot data nodes)
Elasticsearch
(warm data nodes)
Kibana
Logstash
Logstash
Beats agent
Redis Cluster

Consul 架构图


master1
master2
master3
agent
agent
agent
agent
Kibana
Elasticsearch
Logstash
Redis master
kibana.service.consul
master.elastic.service.consul
ingest.elastic.service.consul
data.elastic.service.consul
logstash.service.consul
redis.service.consul

部署前的准备

更新系统

- 如果条件允许，建议把系统软件更新到最新版本，并重启一次。
- 部署过程中，由于系统差别，可能会遇到某些依赖包缺失。请自行修复。离线部署可到 https://pkgs.org/搜索下载。

Elasticsearch 数据盘容量估算公式

- 以每条日志 1kb 大小为例，每秒产生 1000 条日志记录，不做任何解构的前提下，每天的数据存储量在 83GB 左右。 2TB 磁盘空间约可以存储 23 天左右的日志。数据盘必须独立加载，不与系统盘共享空间。其它硬件要求看下一页。
  *1Kb*1000*86400≈83GB*
- 经过 Logstash 解构过滤后的日志存储量，会比实际的低。
- 分布式架构下，每个 ES 数据节点的数据盘存储使用率，应限制在 85% 以下，保留足够的冗余空间，预防个别节点故障时的 index 存储迁移。

配置要求

名称	要求
操作系统	Ubuntu xenial （首选），次选 CentOS 7 。
CPU	4U+ ， 2.6GHz+ 。
RAM	8GB+(如果是单点部署 ELK 三个应用，不低于 16GB 。）
物理数据盘	非系统分区，独占。分区格式 XFS ，容量根据计算公式推导。
平台	建议使用云主机，便于资源扩容。

压测配置

数据来自： https://elasticsearch-benchmarks.elastic.co

CPU: Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz

RAM: 32 GB

SSD: Samsung MZ7LN512HMJP-00000

OS: Linux kernel version 4.13.0-38

OS TUNING:
/sys/kernel/mm/transparent_hugepage/enabled = always
/sys/kernel/mm/transparent_hugepage/defrag = always

JVM: Oracle JDK 1.8.0_131-b11

压测结果（3 节点）

defaults: out of the box configuration of Elasticsearch

4g: 4GB heap size

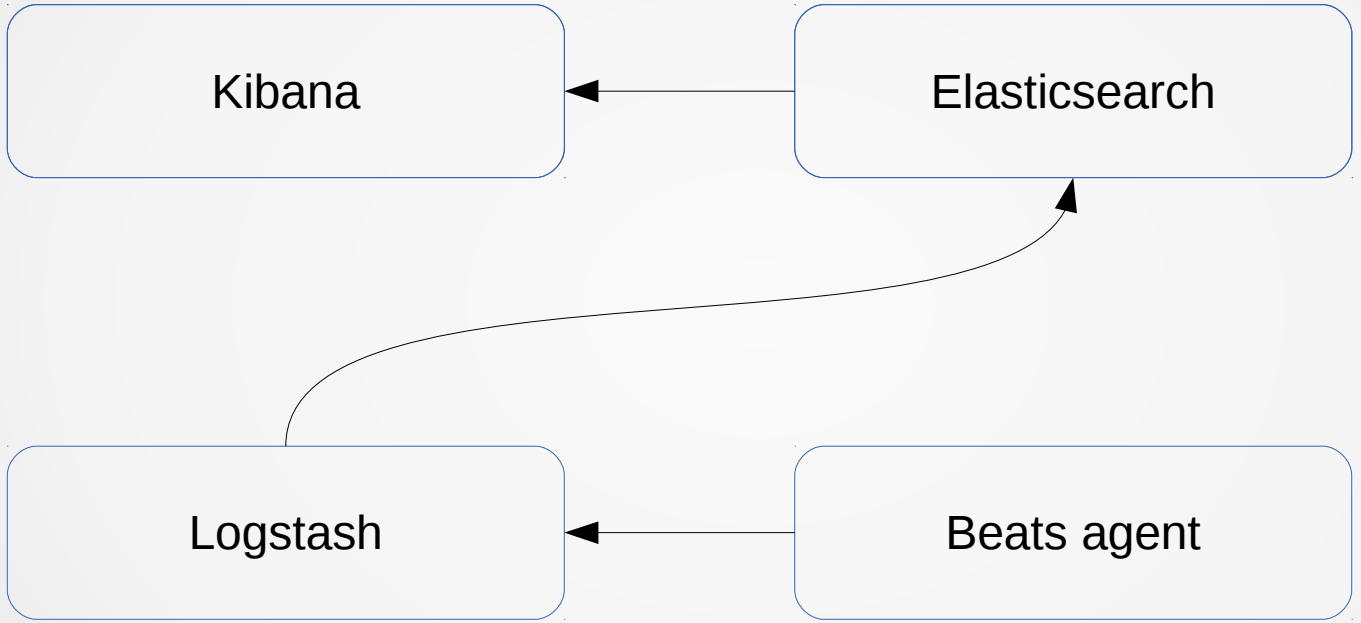
security: with X-Pack Security enabled

3-nodes: runs against a three node cluster (with one replica)

no-src: The _source field is disabled

HTTP Logs	日志数	数据大小
Indexing Throughput	约 12 ～ 18 万条 / 秒	16GB/24 小时
Io Index Size (src on)	(同上)	34.3GB/24 小时
Io Written (src on)	(同上)	112.4GB/24 小时
Io Index Size (src off)	(同上)	26.6GB/24 小时
Io Written (src off)	(同上)	88.1GB/24 小时

单点部署架构（仅开发测试）

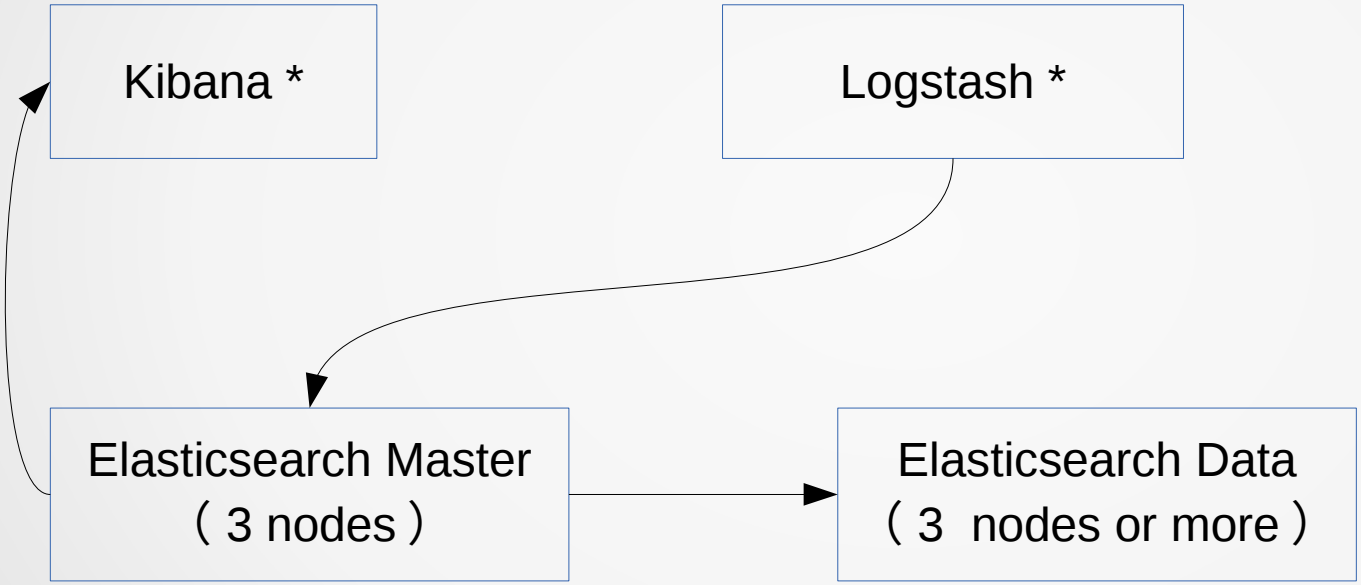


小型部署架构（3 节点）

应用场景：对数据完整性以及集群高可用有要求，没有实时弹性伸缩要求。

中型部署架构（弹性扩缩）

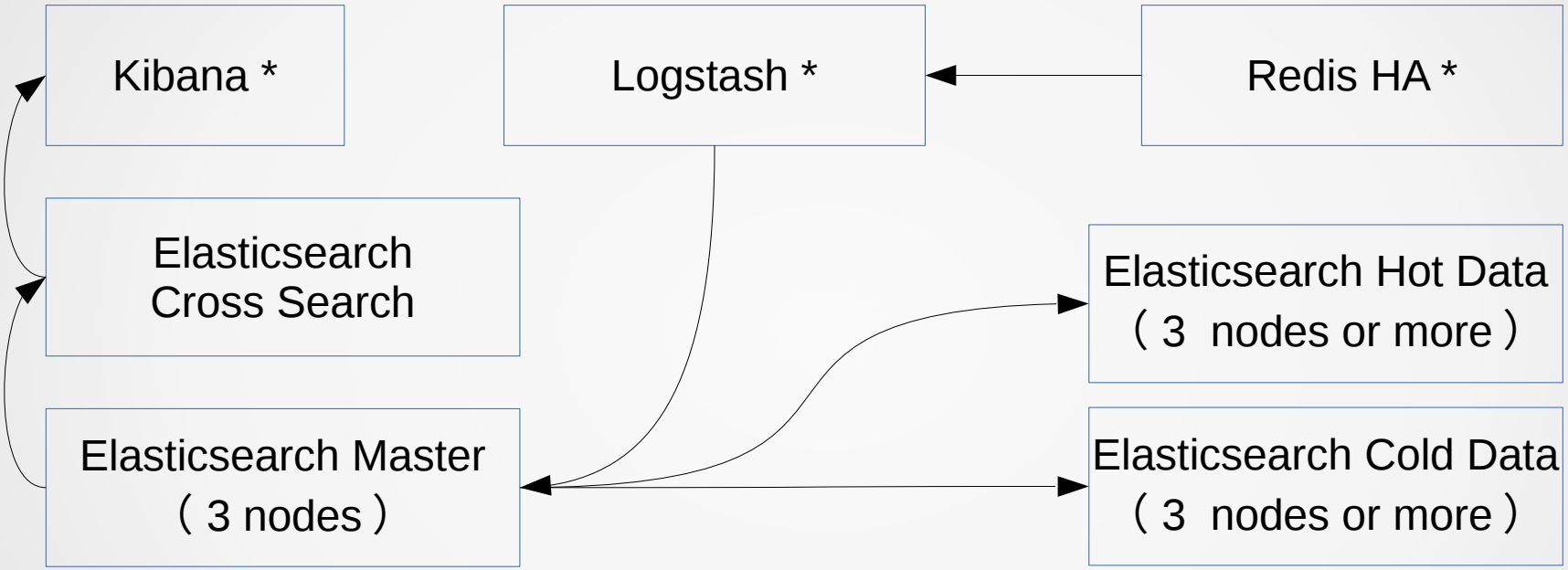
应用场景：业务量会不断增长，需要保留长期历史日志，但对实施搜索要求不高的。



- 带 * 表示至少 1 个节点并可以扩缩。
- Elasticsearch Data 节点可扩缩。
- Elasticsearch Master 负责维护集群节点信息，以及处理 index 。不存储数据。

大型部署架构（弹性扩缩，冷热分离）

应用场景：日志流量高低峰波动大，跨机房异地高可用，对实时搜索有要求。



- 带 * 表示至少 1 个节点并可以扩缩。
- Redis 作为前置缓存，大内存应对日志高峰，避免日志采集端堆积数据。
- Elasticsearch Hot Data 节点可扩缩，使用 SSD ，存储最新数据。
- Elasticsearch Cold Data 节点可扩缩，使用大容量机械硬盘，存储历史数据。
- Elasticsearch Master 负责维护集群节点信息，以及处理 index 。不存储数据。
- Cross Search 节点可以跨多个 ES 集群进行查询，本身不处理数据。

数据盘空间管理

- 使用官方 curator，通过定时任务把旧数据删除或迁移到冷数据节点。
- 通过 zabbix 监控，超过使用率上限时，强制清理数据或迁移。
- 1 台 Data 节点可挂载多个数据盘，在 elasticsearch.yml 配置里添加对应路径即可。 Elasticsearch 集群会自动平衡数据的分布。
- 也可以通过添加 Data 节点来增加可用空间， Elasticsearch 集群会自动平衡数据的分布。
- 如果使用分布式存储，可将索引的副本数量设为 0, 节省磁盘空间和 IO 吞吐。

常见问题

- Q： 如何升级?

  A： 根据官方的升级手册，通过新致提供的 ansible 自动化脚本，实现滚动升级。升级前可对数据做快照，对数据无害。

- Q： ELK 的访问限制?

  A： 方法一是购买官方的 X-PACK 服务，最为理想；

  方法二是通过 nginx 反向代理，添加 http(s) 登录验证，维护稍微麻烦；

  方法三是使用多个 ES 集群隔离不同的日志，并设置 http(s) 登录验证。

  方法二 / 三均不能对 ES 数据做保护，例如删除数据。