在线es统一建设方案

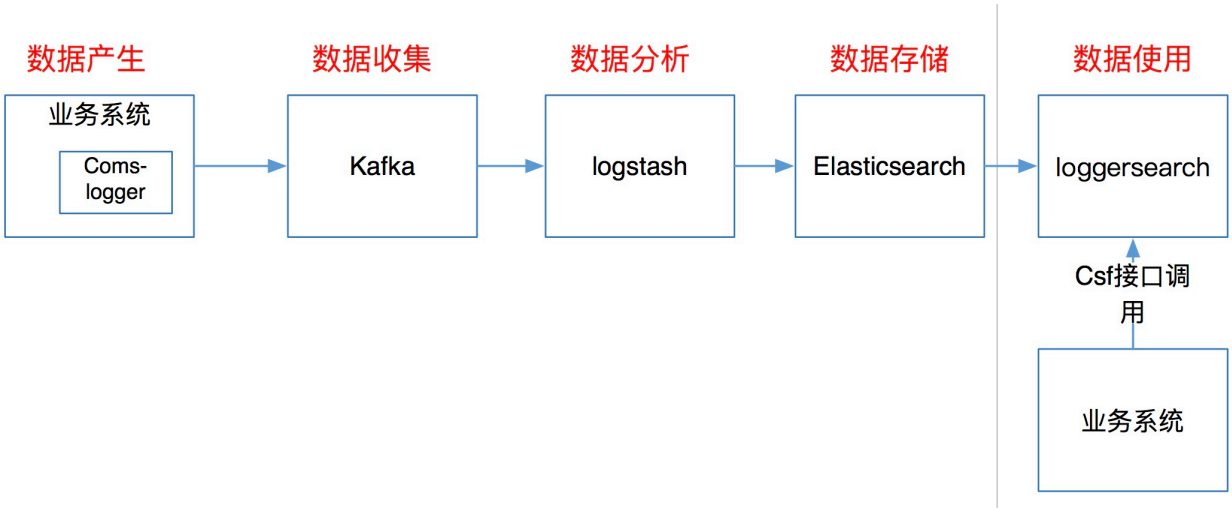
背景

ElasticSearch作为高扩展分布式搜索引擎，主要满足于海量数据实时存储与检索、全文检索与复查查询、统计分析。在如今大数据时代已经成为较popular的存储选择。由于传统mysql不支持大字段、海量数据全文检索、复杂查询，es搜索引擎在中移在线越来越受欢迎。

目前在线es需求一部分使用了平台组统一搭建的es平台，另一部分由业务系统自行部署和维护。自行维护的es从架构设计-资源预估-部署规划-使用规范-运行维护，都没有统一的规范和标准。这就导致：1.资源浪费、es集群性能低 2.违规查询导致es集群频频down机 3.版本、配置各种各样，缺少统一监控组件，排查问题时两眼一抹黑。

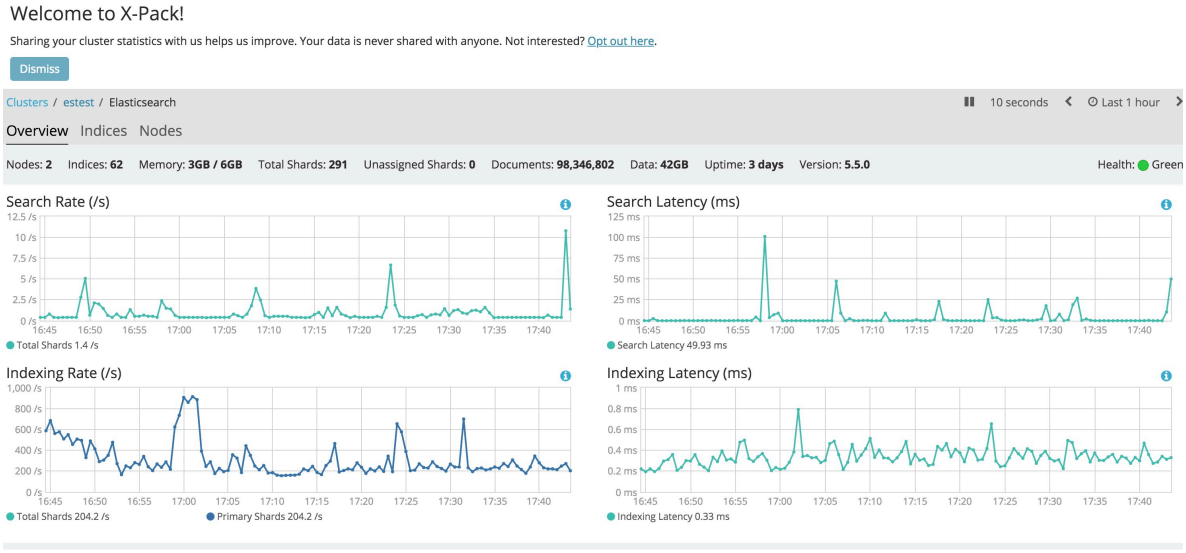
计划对es搜索引擎进行统一资源申请、部署搭建和维护，es在使用时制定良好的规范和约束，以服务的形式提供给业务系统。这样一方面减轻运维复杂度、增加es能力沉淀，另一方面减少业务系统使用搜索引擎的工作量、更专注于业务开发。

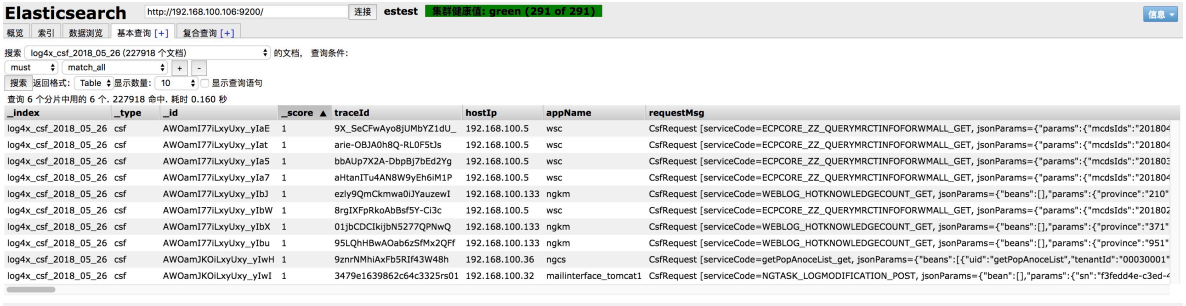
技术架构

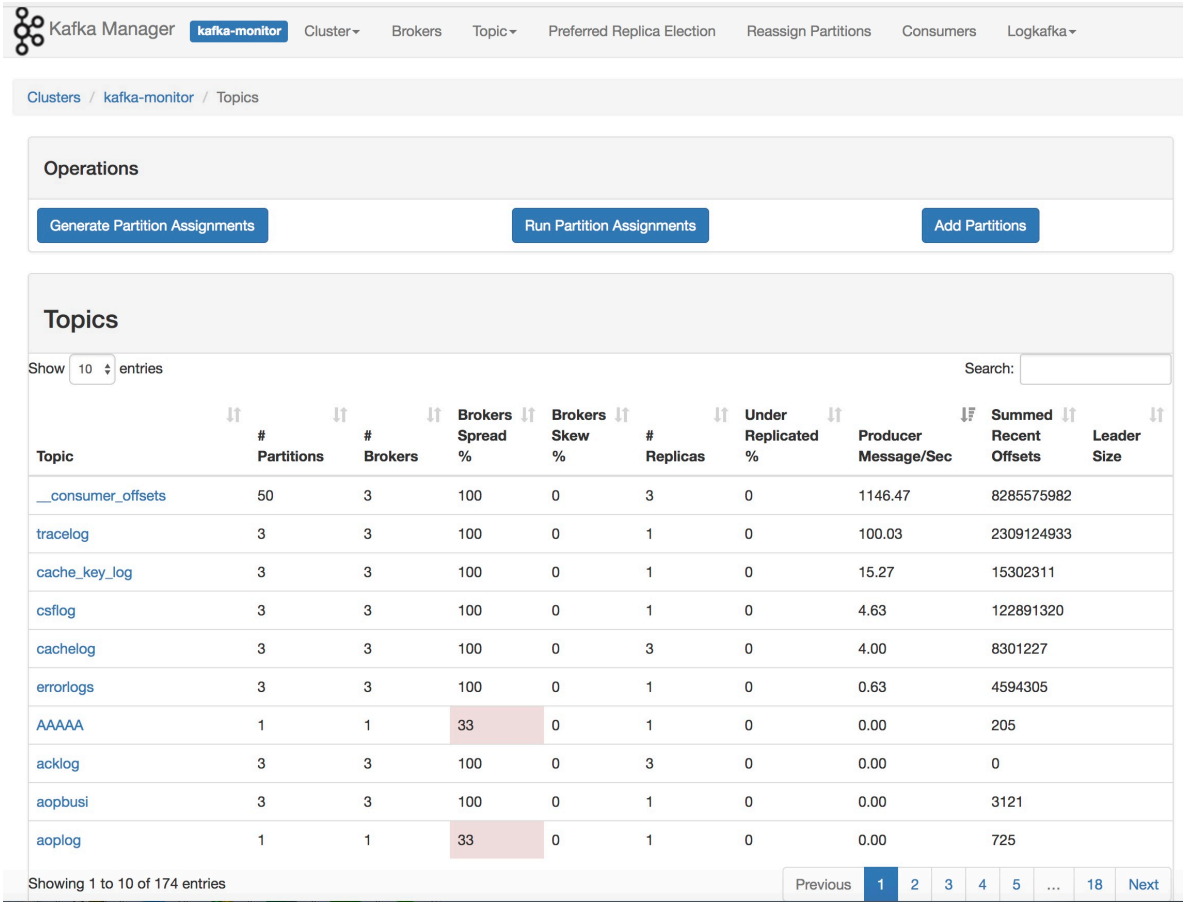


1. 业务系统写数据时：首先业务系统调用日志客户端接口将数据异步写入kafka集群，由于日志客户端内部使用了阻塞队列和kafka故障切换机制，保障写数据操作完全异步，kafka故障也不会影响业务运行。然后logstash实时消费kafka中数据，按照预先定义好的规则对数据进行清洗、过滤，最后将数据持久化到es中。每类数据对应kafka中单独topic、单独logstash、es中单独索引，不同类别数据的存储流程相互独立。
2. 业务系统查询数据时：不允许业务系统直接连接es集群进行操作，而是调用由loggersearch（自维护java工程）暴露出的csf接口，我们在loggersearch中做了租户和查询条件安全限制，同时对csf接口增加了失败率和超时监控，保障所有查询操作都在控制范围内。当es集群故障或升级时，通过配置中心的配置热修改将loggersearch提供的csf服务快速失败，同时在csf响应报文中塞入特定状态码，业务系统接受到特定状态码后可做降级处理。

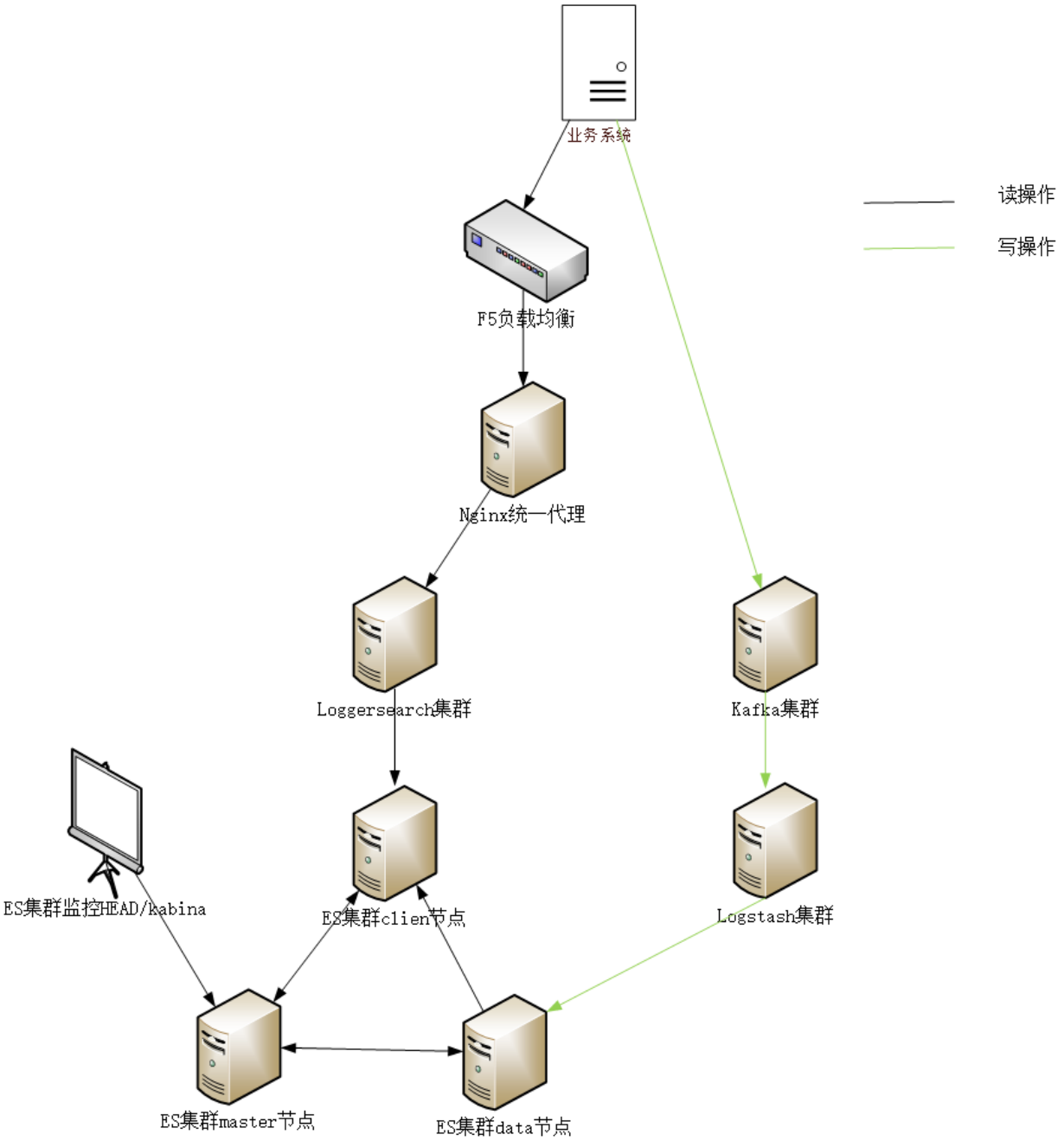
3. 监控维护方面：使用kafka-manager、es-head、x-pack和kinaba对kafka、logstash、es组件进行资源和性能监控。此外所有数据都需要做生命周期管理，使用调度平台定时调用loggersearch的索引关闭和删除服务，保障es集群健康运行。







部署规划



- **Kafka**：按照3节点点进行建设，需求量较大的按5个节点建设，其余3节点。使用虚拟机、允许交叉部署，该机器需要分配至少1T存储
- **Logstash**：一个需求对应一个logstash，每套集群预留两台机器部署logstash。使用虚拟机、允许交叉部署，数据量超过1W/s的需求logstash单独部署，该机器需要cpu、内存较高
- **Elasticsearch**：按照master节点、client节点、data节点分别建设，每类节点的规范和资源要求不同
    - **master节点**：主要用于元数据(metadata)的处理，比如索引的新增、删除、分片分配等。按照3节点建设，使用虚拟机、允许交叉部署，jvm内存2-10G
    - **client节点**：主要用于接受和分发检索请求，可当成负载均衡节点。按照3节点建设，使用虚拟机、允许交叉部署，jvm内存2-10G。增加client节点可增加检索并发,但检索的速度还是取决于查询所命中的分片个数以及分片中的数据量
    - **data节点**：主要存储数据、接受索引请求，所有的检索最终都落到该节点上。按照5节点建

设，使用虚拟机、单独部署，机器要求：4-8C、64内存、1-2T存储

- **loggersearch**：暴露csf接口的应用，按照5节点建设，使用虚拟机、允许交叉部署，无资源要求

建设规划

计划全部进程部署在虚拟机上，便于后续迁入pass平台。按照业务团队划分es集群，计划按照一体化、智能化、网络天下、互联网团队、对外业务拓展中心、平台组（csf、redis、gis等）搭建6套es平台。按照分批次、最小规模建设，单个集群最小规模资源规划如下：

统一ES服务器资源需求（单个集群最小规模）						
类型	服务器	作用	最低配置	数量	部署节点数	备注
kafka+zk	云平台虚拟机	消息队列用于写数据的缓冲	CPU：4核 内存：16G 存储：1T	3	3*2	根据业务访问量做节点调整。
loggersearch-web、service	云平台虚拟机	对外暴露csf接口，用于控制和记录查询操作			3*2	
logstash	云平台虚拟机	分析端用于消费kafka数据写es	CPU：8核 内存：64G, 存储：200G	2	最大16	根据业务访问量做节点调整。
ES client节点	云平台虚拟机	es负载均衡节点，接收查询请求	CPU：4核 内存：16G, 存储：200G	3	3	根据业务访问量做节点调整。
ES master节点+可视化ui	云平台虚拟机	es主节点，管理集群元信息			3+3	
ES data节点	云平台虚拟机	es数据节点，存储数据、执行查询、聚合操作	CPU：8核 内存：64G, 存储：2T	5	5	根据业务访问量做节点调整。
合计				13		

建设排期：第一批6月15号完成一体化和智能化es平台部署、第二批6月22号完成互联网团队和平台组es平台部署、第三批6月29号完成对外业务拓展中心和网罗天下es平台部署

迁移与维护

迁移排期：荆培洋负责推动接入平台组es的业务需求迁移到统一es平台中，郭拥宾负责推动业务系统单独部署的es迁移到统一es平台中

维护：ES集群由运维组统一搭建管理，各业务团队根据自己业务需求向运维组申请ES集群，后期维护和升级工作也由运维组统一完成，原则上不再为业务团队单独申请主机搭建ES集群。维护工作划分如下:

1. 业务系统接入es前的需求评审（平台组+运维组）
2. 根据业务侧反馈的数据结构，编写索引模板和logstash配置（运维组）
3. 在多个环境创建topic、logstash、索引、定时调度（运维组）
4. 上线前数据迁移 （运维组）
5. 数据、集群异常问题排查（运维组）
6. 集群升级方案、性能优化方案、功能新增、难点公关（平台组）

大数据es运营流程

整个流程结束，会沉淀以下文档：1.需求文档 2.数据结构和索引模板 3.logstash配置 4.数据定期清除规则