



# Day4 三步搭建自己的google搜索引擎

# 无处不在的搜索

搜索无处不在，尤其是在移动互联的今天。无论是社交，电商，还是视频等APP中，搜索都已经在其中扮演了重要的角色。作为信息的入口，搜索能帮用户从海量信息中找到想要的信息。

在搜索框中快速依据关键字，找出对应内容，这是网站或者APP不可或缺的一个功能。搜索性能、精准度，又是和最终用户使用体验息息相关的。

[云搜索服务](#)正是在**高性能、全文检索、自定义词库、多种类型分词、模糊查询、自动纠错、同近义词匹配、简繁体匹配**等方面，来满足用户对搜索的诉求。



# 场景描述

Google搜索引擎大家都用过，Google的搜索内容，都是依靠爬虫从网络上爬取的网页数据。当用户在输入框中输入关键字，Google就会在后台进行内容相关性计算，然后以相关性从高到低返回结果。

对于一个网站或者APP来说，我们也希望拥有类似于Google的搜索框。只不过搜索的内容，是网站自己提供的内容。

本实战场景，就以一个网站的搜索框作为举例，这个网站的内容包含注册用户，以及用户发的帖子。通过构建网站的搜索框，带您体验[网站搜索场景](#)中云搜索服务的应用。其中包括三个搜索场景：

- ❑ 用户搜索，根据用户名称匹配度、是否VIP、最近登录时间、发帖数、粉丝数综合评分得出搜索结果
- ❑ 帖子搜索，根据帖子内容匹配度、发布时间、查看数、评论数、点赞数综合评分得出搜索结果
- ❑ 同义词搜索，根据表达意思相同的词，得到相应的搜索结果

# 场景理解

网站中的用户数据:

```
{
  "username": "刘德华",
  "is_vip": true,
  "postNum": 100,
  "fansNum": 1988,
  "lastLogin": "2018-07-04 12:01:12"
}
```

注册用户名  
是否VIP用户  
发帖数量  
粉丝数量  
最后一次登录时间

网站中的帖子数据:

```
{
  "desc": "华为云搜索服务助力企业解决部署周期长、运维难、使用成本高等问题，让用户更专注于自身业务发展。",
  "playNum": 46344,
  "commentNum": 233,
  "likeNum": 6,
  "createTime": "2018-07-17 09:00:00"
}
```

发帖内容  
浏览数量  
评论数量  
点赞数量  
发表时间

实现效果:

搜索框中输入关键字，可以从用户或者帖子中搜索相关内容

Web Search Demo (HWS\_MyName)

用户搜索 帖子搜索

刘德华

用户名	是否VIP	最近登录时间	发帖数	粉丝数	相关性评分
刘德华	true	2018-07-04 12:01:12	100	1988	108.15413
刘德华忠粉	true	2018-06-05 09:01:00	100	1225	78.514206
刘德华5	false	2018-07-11 08:01:34	475	233	66.01069
刘德华将军	false	2018-05-04 12:01:12	48	566	48.66796
邵阳刘德华	false	2018-07-15 12:56:12	10	21	27.525759
刘德华	false	2018-07-12 11:01:00	10	188	22.56369
刘德华	true	2018-06-04 12:01:32	4	21	8.563893
翻唱刘德华	false	2018-05-04 12:01:12	2	0	5.62239

Showing 1 to 8 of 10 rows 8 rows per page

Web Search Demo (HWS\_MyName)

用户搜索 帖子搜索

华为云搜索

帖子内容	发布时间	查看数	评论数	点赞数	相关性评分
华为云搜索服务助力企业解决部署周期长、运维难、使用成本高等问题，让用户更专注于自身业务发展。	2018-07-17 09:00:00	46344	233	6	57.90577
华为云上的云搜索服务是将开源软件 ( Elasticsearch, Kibana ) 及硬件(计算、存储、网络)云化后，为用户提供的全托管一体化的搜索和分析平台	2018-07-21 06:00:00	334	23	1	52.360085

Showing 1 to 2 of 2 rows 8 rows per page



# 搜索框构建步骤



# 准备环境

## 云搜索服务(Cloud Search Service)

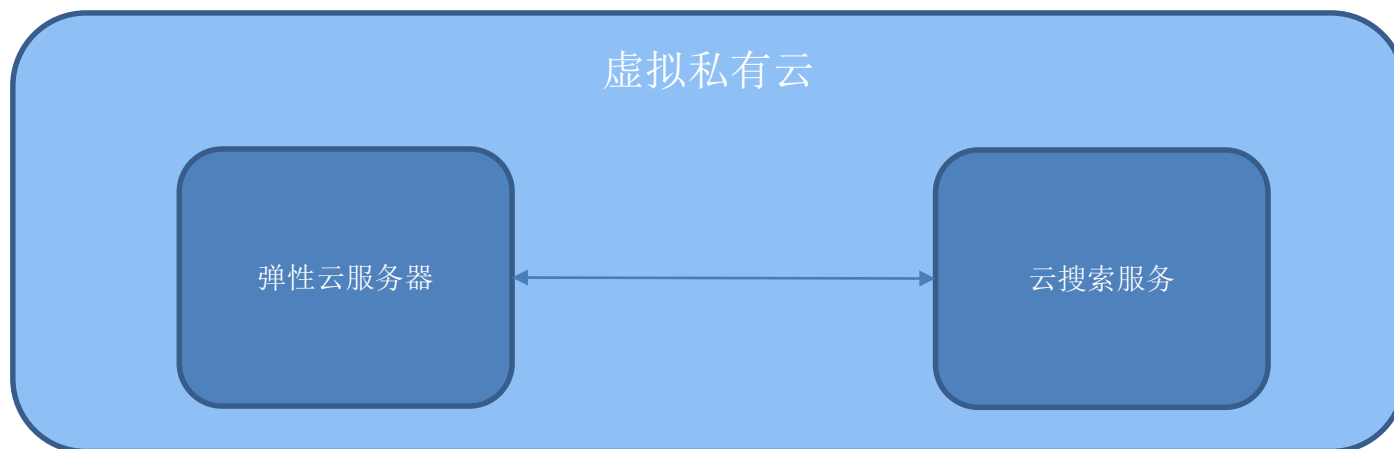
是一个基于Elasticsearch且完全托管的在线分布式搜索服务，为用户提供结构化、非结构化文本的多条件检索、统计、报表。完全兼容开源Elasticsearch软件原生接口。

## 弹性云服务器（Elastic Cloud Server）

是一种可随时自助获取、可弹性伸缩的云服务器，帮助用户打造可靠、安全、灵活、高效的应用环境，确保服务持久稳定运行，提升运维效率。

## 虚拟私有云（Virtual Private Cloud）

是用户在华为云上申请的隔离的、私密的虚拟网络环境。用户可配置VPC内的IP地址段、子网、安全组等子服务，可申请弹性带宽和弹性IP搭建业务系统。



# 索引数据

## IndexingApp的功能

索引数据是将数据源中的数据，在Elasticsearch集群中建立索引。通常会使用数据同步工具或者ETL工具，从MySQL拉取数据，写入Elasticsearch。

在本次课程里，我们的数据源是磁盘上的很多JSON文件，因此，我们写了一个程序IndexingApp，从磁盘上读取文件内容后，调用Elasticsearch的写接口，将输入写入Elasticsearch集群。

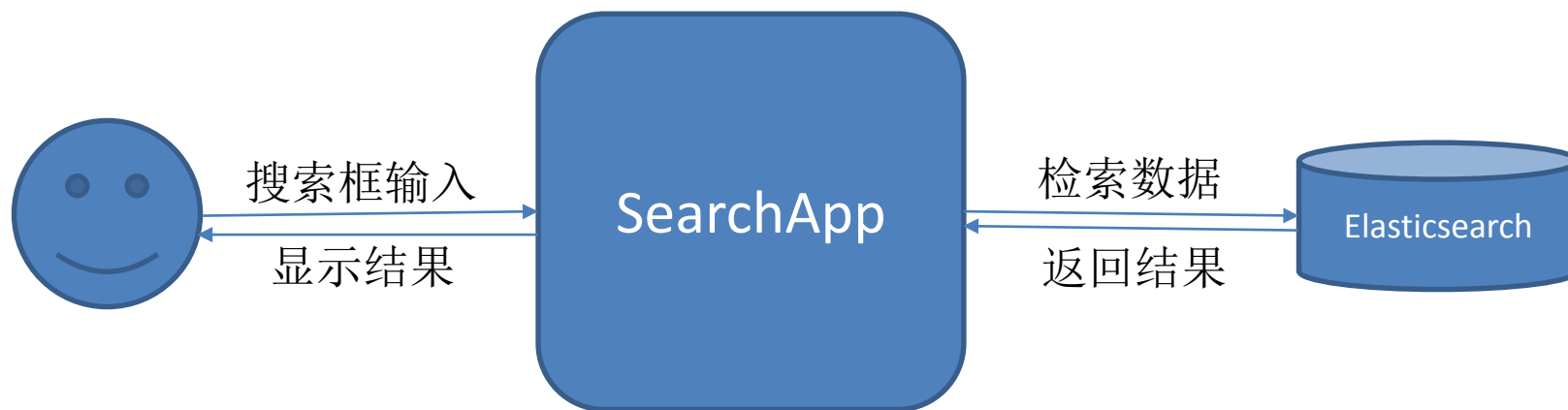


# 搜索请求

## SearchApp的功能

搜索请求是从搜索框中接收用户的输入，然后构造搜索查询语句，调用Elasticsearch的搜索接口下发检索数据请求到Elasticsearch。

在本次课程里，我们试用SearchApp来模拟一个网站的前后台。结合用户输入和搜索范围选择，来显示相应的搜索结果。





# 精彩预告

本节课，我们搭建了一个网站搜索框，体验了云搜索服务怎么和应用程序进行交互，以及同义词搜索功能的实现。

下节课，我们会基于云搜索服务中的Kibana，来看看怎么做出图形可视化页面。



# Thank You.

**Copyright©2018 Huawei Technologies Co., Ltd. All Rights Reserved.**

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.