

开放搜索

产品简介

产品简介

产品概述

参与《开放搜索产品咨询》问卷，赢取代金券！

OpenSearch产品概述

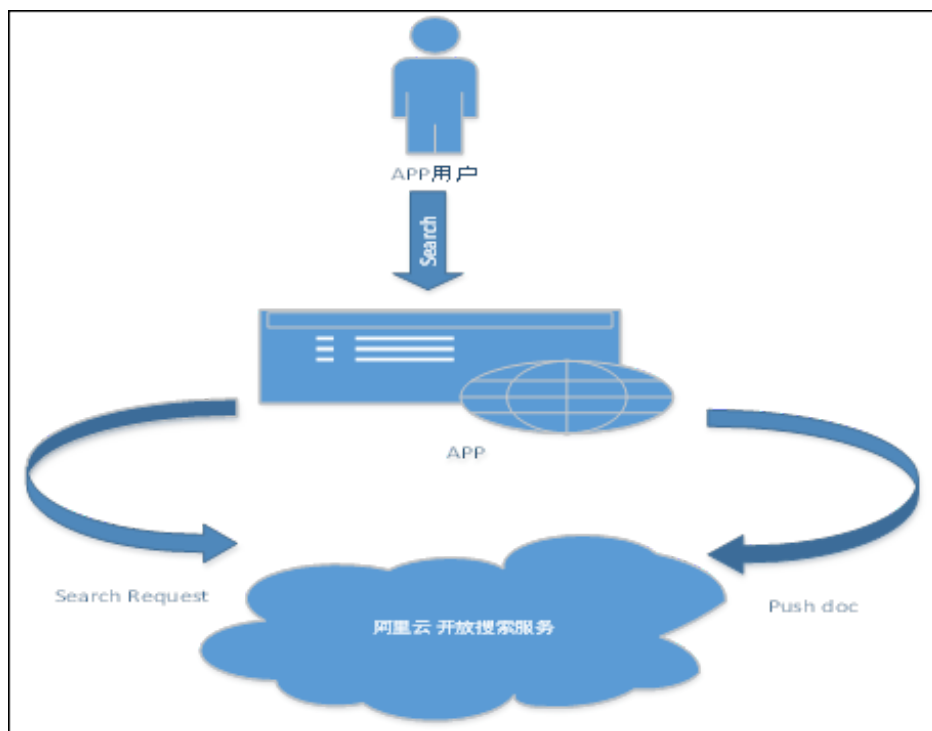
简要介绍

开放搜索（OpenSearch）是一款结构化数据搜索托管服务，为移动应用开发者和网站站长提供简单、高效、稳定、低成本和可扩展的搜索解决方案。

OpenSearch基于阿里巴巴自主研发的大规模分布式搜索引擎平台，该平台承载了阿里巴巴全部主要搜索业务，包括淘宝、天猫、一淘、1688、ICBU、神马搜索等业务。OpenSearch以平台服务化的形式，将专业搜索技术简单化、低门槛化和低成本化，让搜索引擎技术不再成为客户的业务瓶颈，以低成本实现产品搜索功能并快速迭代。

使用OpenSearch搭建搜索服务，您只需：

1. 创建搜索应用
2. 编辑您的应用结构
3. 上传数据
4. 从您的网站或应用程序提交搜索请求



开放存储服务ODPS、RDS用户还可以在OpenSearch控制台直接配置使用相应的数据源，数据将自动同步进入OpenSearch，简单、方便、可靠。OpenSearch后续将支持更多的数据源自动同步，例如OTS等；提供更丰富的搜索外围功能，例如相关搜索、搜索热词等。敬请期待！

为什么选择OpenSearch?

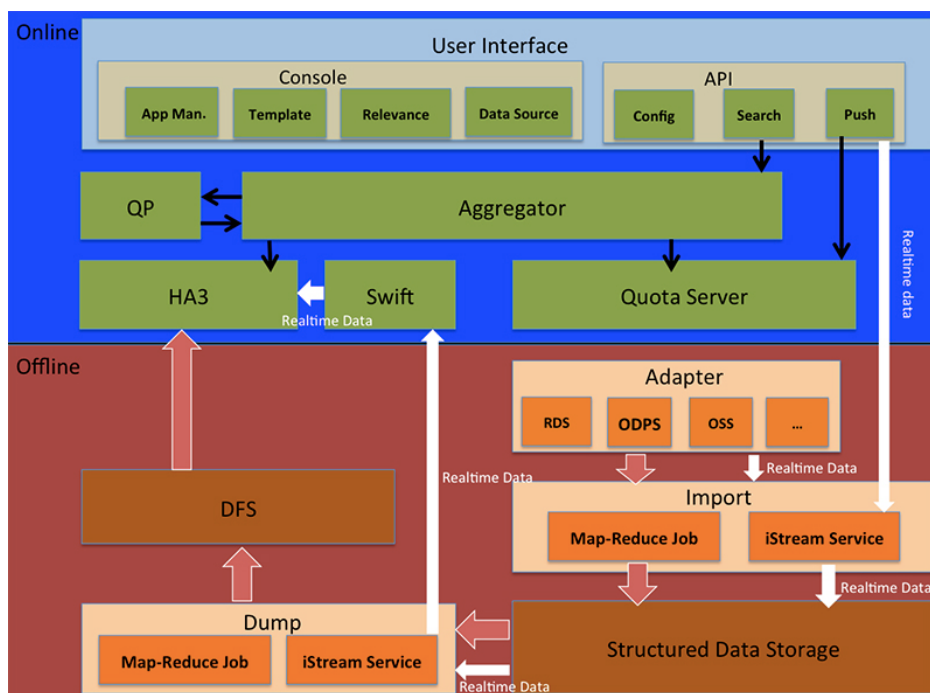
1. 支持用户上传数据或同步云数据，实时性有保障。
2. 应用结构、排序相关性自由定制，搜索服务更个性化。
3. 基于阿里巴巴在搜索领域的积累，提供查询分析功能，对用户查询词进行纠错、词权重分析、停用词过滤，让搜索服务更智能。
4. 可视化的界面、丰富的模板，不用精通代码也能快速创建自己的搜索应用。

选择OpenSearch，就选择了简单、高效、低成本和可扩展的搜索解决方案！

产品架构

实现原理

OpenSearch基础架构



- 白色线为实时数据处理流
- 红色线为全量数据处理流
- 黑色线为搜索流程

开发者通过控制台和API与系统交互。典型的使用流程是开发者进入控制台，创建应用实例，配置应用字段结构、搜索属性，配置文本处理插件、定制相关性排序规则等。应用实例创建完成后，开发者再通过SDK/API将数据推送至云端（阿里云存储用户可以配置数据自动同步，只需在控制台中授权），数据实时流式进入Import子系统的数据库导入服务模块(iStream Service)，经过格式解析和数据处理后，存储在结构化数据存储系统中。随后，Dump子系统的数据库导出服务(iStream Service)将数据经过一定处理后发送给实时消息队列系统(Swift)，搜索系统(HA3)从消息队列中订阅数据，在内存中构建索引并提供搜索服务。这个数据实时流式处理过程（白色箭头）大概十秒左右。

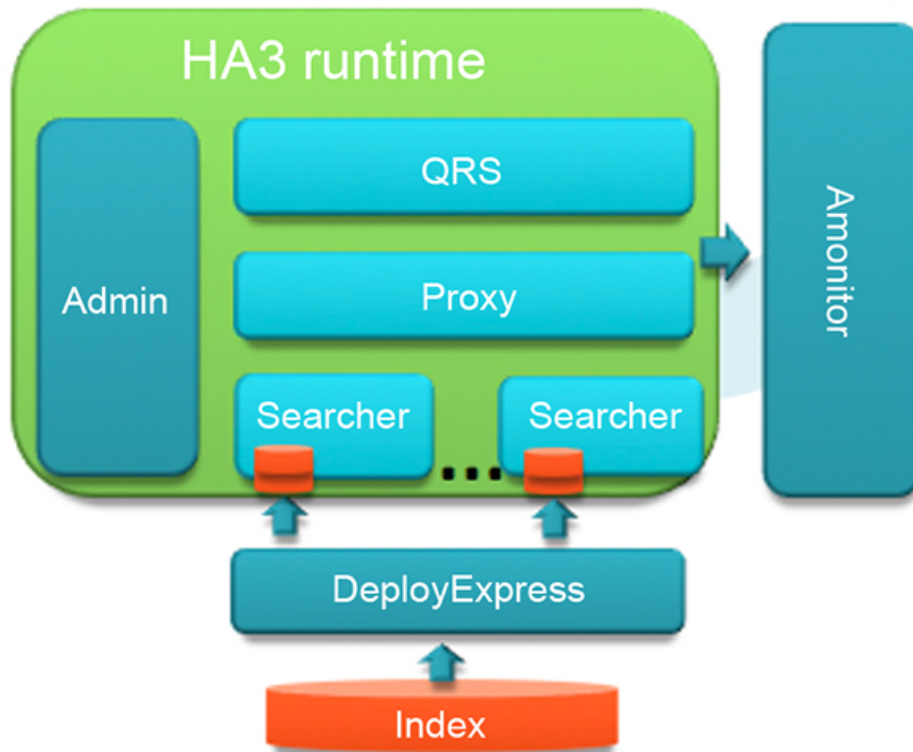
当开发者修改了索引结构，需要对应用中的数据做增量索引重建。为了保证搜索效率，系统也会定期对所有数据做全量重建索引。索引重建流程参见红色箭头，这是一个非实时的流程，依数据大小不同可能需要几分钟到十几分钟，全量索引重建则需要数小时。

数据在云端经过一系列处理和索引构建后，开发者就可以通过API搜索应用实例中的数据。搜索请求首先发送到查询聚合服务Aggregator。如果开发者配置了查询改写处理逻辑（即将上线），Aggregator会将查询请求发送给查询改写服务QP，QP按照开发者配置的处理规则（例如：拼写纠错、同义词或者查询语义改写）改写查询请求，并将改写后的查询回传给Aggregator，Aggregator最终将查询请求发送给搜索系统HA3，HA3根据开发者定制的相关性排序规则对命中的结果文档排序，并最终通过Aggregator将结果返回给开发者。

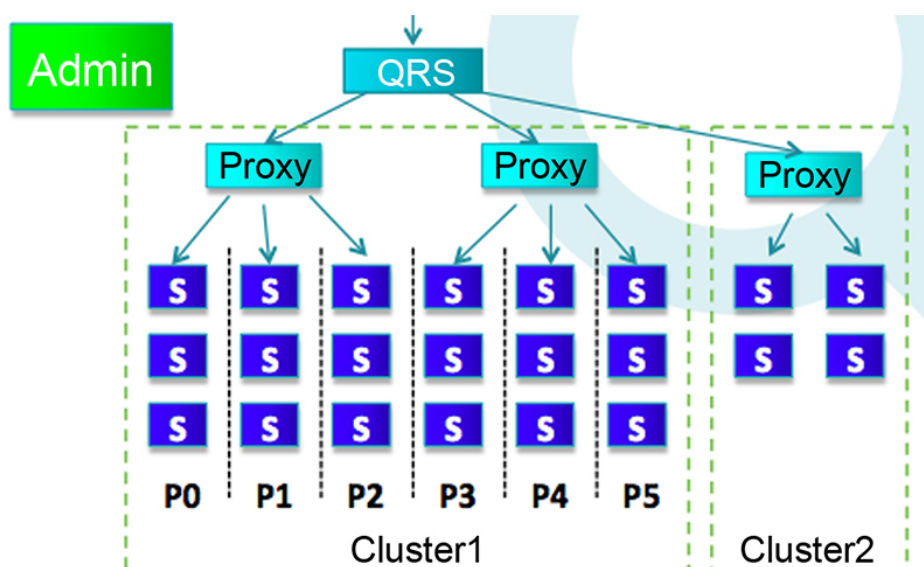
为了保证不同开发者各个应用数据推送和搜索相互不受影响，资源合理利用。配额管理服务（Quota Server）会对进入系统的数据和搜索请求频率依据开发者的配额（文档总量、QPS）做限流控制。超出配额部分的数据推送将失败，查询请求将随机丢弃。

引擎实现原理

前面多次提到的HA3是阿里自主研发的新一代分布式实时搜索系统，中文名叫问天3，具备自动容灾、动态扩容、秒级实时等能力。下图是HA3系统模块组成图。



其中，Admin是整个系统的大脑，负责节点角色分配、调度决策、FailOver处理、状态监测、动态扩容等。Amonitor是系统的性能状态监控模块，收集和展示整个系统所有节点的性能参数。QRS是查询解析和改写服务，是系统对外的搜索接口。Proxy是搜索代理模块，负责接收QRS的查询请求，并转发给下辖的所有Searcher节点。Searcher节点执行实际的查询匹配计算，将搜索结果汇总后回传给QRS。DeployExpress是分布式链式数据实时分发系统，负责将离线集群构建好的索引数据分发到各个Searcher节点。DeployExpress的最大亮点是将1份数据分发多份拷贝到Searcher节点，其分发时间接近单份拷贝的数据分发时间，而且单节点故障能自动恢复，不影响数据拷贝。在同等硬件条件下，基于1200万数据做单机性能对比测试发现，HA3比ElasticSearch开源系统的QPS高4倍，查询延迟低4倍。



上图是HA3的多集群异构部署图，其中部署了两个异构逻辑集群Cluster1和Cluster2，两者的硬件配置、索引结构、服务能力可以不同。这种部署一般用来实现冷热数据分层查询、异构数据查询等功能。

OpenSearch利用异构逻辑集群优化资源配置，提升系统服务能力和降低机器成本。不同特性的应用实例被分配在不同的逻辑Cluster中。例如，QPS较高，数据量较少的应用实例分配在SSD磁盘的Cluster中，该Cluster列数较少，但行数较多，能承载较大的搜索流量；而一些QPS较低，数据量又较大的应用实例分配在普通磁盘Cluster中，该Cluster行数较少，但列数较多，能承载海量的用户数据。当每个逻辑集群的数据量增大时，系统可以通过增加列（Partition）来扩大系统数据容量；当搜索流量增大时，通过增加行（Replicas）来提升系统服务能力。

功能特性

OpenSearch有以下一些主要功能。

支持文档索引结构定制，以及自由修改

OpenSearch将搜索引擎复杂的索引结构概念简单化、可视化和自助定制化。开发者可以通过控制台创建搜索应用，定制文档字段的结构和属性，包括字段名称、类型、分词方式、搜索属性等。搜索应用在运行过程中可以自由修改，满足了产品快速变化的需求，极大缩短了需求变更到上线的过程。

支持主流阿里云存储产品的自动对接，数据自动同步更新

开发者的数据如果在阿里云ODPS、RDS等服务上，开发者只需要在OpenSearch控制台中授权，数据就可以自动同步至OpenSearch中，后续数据的更新也可以自动实时同步（ODPS除外）。而且在同一区域中，从云存储同步数据至OpenSearch免收流量费用。数据不在阿里云上的开发者，可以通过RESTful API或者SDK上传数据，小数据量也可以直接在控制台上传。

支持多表数据推送，及字段文本处理和转换

类似于数据库，每个搜索应用可以创建一张或者多张表，每张表的字段上可以内置数据处理插件，对字段内容做文本处理和转换，例如拼音转换、HTML标签剔除、JSON数据解析等，多个表会Join在一起实现联合查询。数据存放在RDS数据库里的开发者，可以用此功能替代数据库全文检索，实现更高的性能和搜索体验。

支持两轮相关性排序定制，简单灵活加速产品效果优化迭代

搜索结果相关性排序是影响用户体验最关键的一环，OpenSearch支持开发者定制两轮相关性排序规则来准确控制搜索结果的排序。第一轮为粗排，从命中的文档集合里海选出相关文档。第二轮为精排，对粗排的结果做更精细筛选，支持任意复杂的表达式和语法。方便开发者能更准确控制排序效果，优化系统性能，提高搜索响应速度。

产品优势

稳定

服务可用性：不低于99.9% 数据持久性：不低于99.9999% 自动故障检测与恢复：提供7×24小时的运行维护，并以在线工单和电话报障等方式提供技术支持，具备完善的故障监控、自动告警、快速定位、快速恢复等一系列故障应急响应机制

安全

阿里云为用户分配AccessKeyId和AccessSecret安全加密对，从OpenSearch访问接口上进行权限控制和隔离，保证用户级别的数据隔离，用户数据安全有保障。数据冗余备份，保证部分机器宕机的情况下，用户数据不会丢失。

大规模、高性能

OpenSearch提供多种规格配置，并具备弹性扩容能力，用户可根据需要自行在线扩展或缩减所使用的应用资源。OpenSearch使用高端服务器来保障每个应用都拥有良好的性能，查询延迟达到毫秒级别。

简单、低成本

开发者无需理解搜索引擎实现细节，几个简单步骤即可拥有专属搜索服务。内置应用场景模板，资源按需使用，免除开发运维成本。

强大的定制功能

支持应用结构、数据处理、查询分析、以及搜索结果两阶段排序定制，强大的功能，灵活的定制机制，能满足开发者复杂的搜索需求，加速产品迭代。

丰富的外围功能（持续开发中）

支持搜索热词、下拉提示、相关搜索等等一系列搜索外围功能，方便用户展示及分析。

使用场景

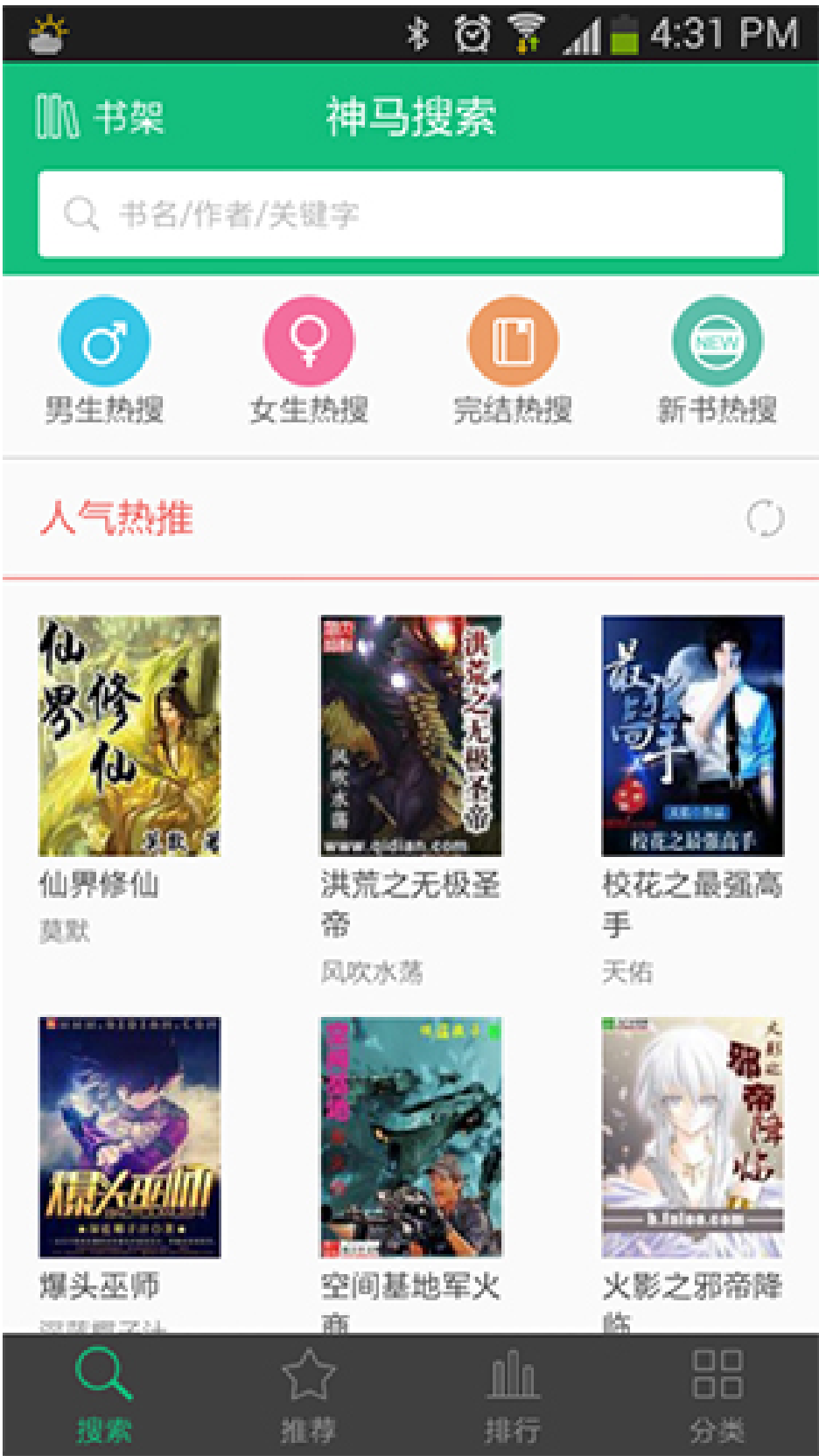
竞品分析

产品测评

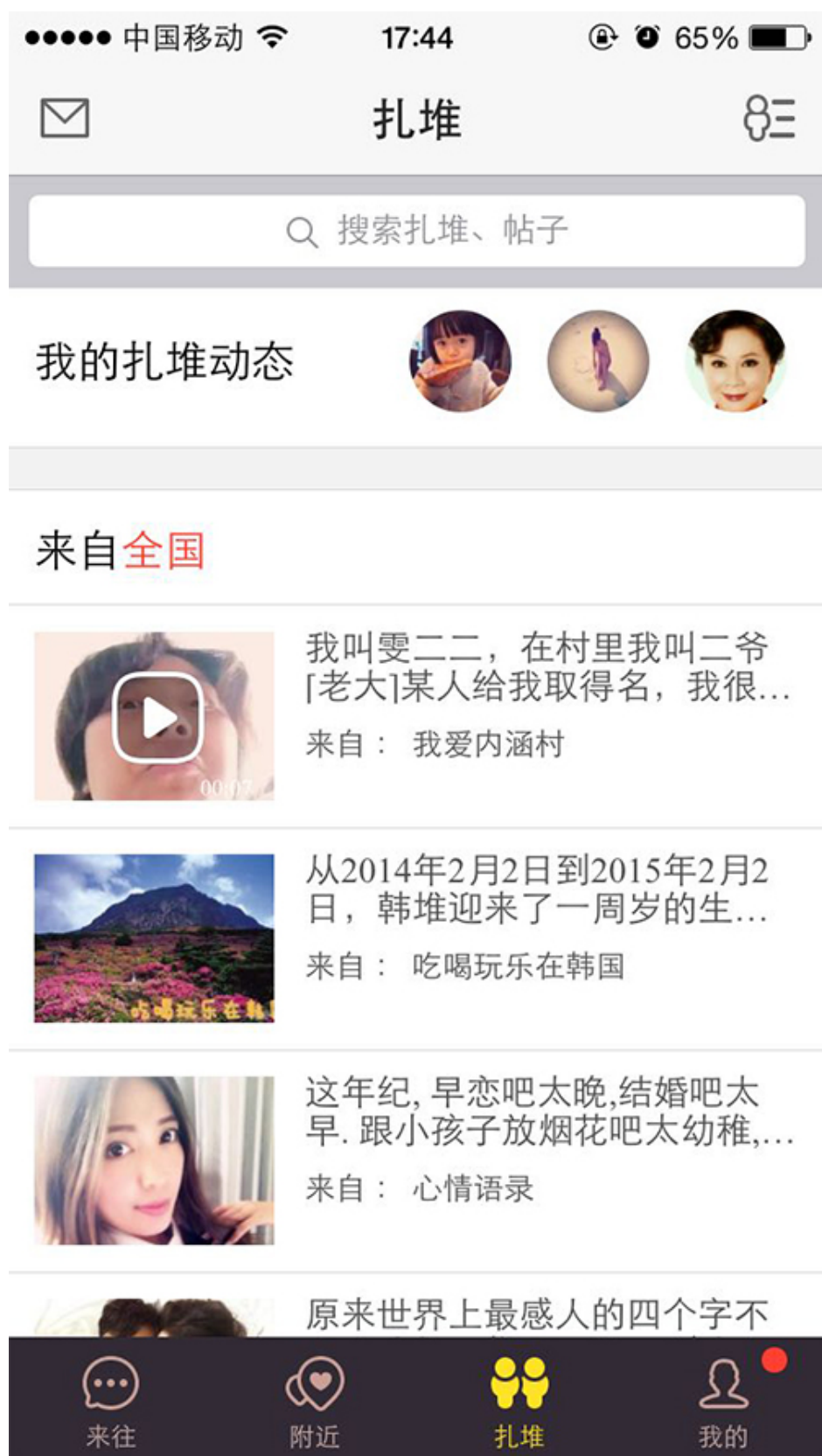
部分用户发表的测评数据，欢迎各位继续补充，我们会定期选择优秀的测评文章进行各个渠道的宣传和推广。[点此进入](#)。

客户案例展示

APP类：神马小说、来往







中国移动

17:44

65%

<

扎堆搜索

🔍 何以笙箫默

✕

取消

扎堆(18)



何以笙箫默

人数: 8.8万 分享: 6907 热度: 96°C

>

查看更多扎堆结果

扎堆文章(262)

有谁跟我一样是看了何以笙箫默才下载的来往? 可惜一个好友都没有😞😞😞

🖼️

Ashl... 分享于 何以笙箫默

评论:312

EXO TAO客串电影版《何以笙箫默》的少年何以琛 据韩媒报道, SM相关人士称EXO黄子韬出演电影版《何以笙箫默》。黄子韬...

🖼️

baob... 分享于 吃喝玩乐在韩国

评论:208

看何以笙箫默才下载的这个软件的小同学举个手手吧🤩

🖼️

社区类: 宝宝树

11



威锋网

威锋网

FENG.COM

ios8字体

全部

论坛

新闻

锋科技

评测

威智网

安锋网

威锋游戏

搜索热词:

ios8字体

ios8美化

tage

主题

字体

ios8主题

auxo3

flex2

8.1.2 越狱

微信多开

8.1.2 越狱教程

iphone6

越狱

Quickdo

ios8 字体

微信 共存

香港

澳门

搜索结果

2626

iPhone 4区精华索引贴第二版 (1月16日更新)

2011-11-12

53810

自游

iPhone 4 综合讨论区

iphone4 想更换系统字体的朋友, 进来看看吧! (感谢作者iviu0) 【1月15日更新版本号面板, 1月7日更新自动生成sua ve图标】 iOS5.0.1全方位手动打造透明主题详细教程 (感谢作者songjr) 4.2.1-4...

『WEIP源小组』【gkdou1989】威锋源所有DEB简介整合总汇『十月...

2011-10-04

615

gkdou1989

iPhone源组 软件讨论区

威锋源现在相对应的分类有: 应用工具、网络冲浪、游戏娱乐、生活应用、系统工具、书籍阅读、增强补丁、键盘输入、越狱破解、汉化补丁、系统字体、开发...

Google思源黑体(Google Noto Fonts)中文字体For iOS8字体替换

一周内

66

lenhoo

iPhone 6 综合讨论区

本帖最后由 lenhoo 于 2015-1-31 02:24 编辑

ios8字体制作教程有吗

2015-01-01

5

糖鸡屎

iPhone 6 综合讨论区

哪位大神有ios8字体制作教程呢 qq里自研的汉仪南宫体很不错 我下到了7.1可以用的 但用在8上不完美 想自己修改 有没有人给个教程 谢谢

ios8字体来咯。喜欢的进来看下。已经更新

2014-12-02

81

蔡溪的喵

iPhone 6 综合讨论区

IOS8字体, 全部都是弄好的, 直接替换就OK了。如果出错方块或者问号, 那就在手机里设置里关掉那个选项, 重启就可以了。这些字体我全部已经自己测试过了。我的是8.1.1版本。下面的字体都...

【不定期更新iOS8字体】系统DEB越狱字体安装包放出★

2014-11-23

189

duguyikuan

iPhone 5s 综合讨论区

注意字体的兼容性, iOS8的字体不能返过来安装到7的设备上, 会很高几率出现乱码。有意思的是部分7.1.x的字体兼容iOS8。注意心态的稳定性, 字体包安装结束重启手机时, 偶尔有可能会遇到假死...

搜索排序

相关度

发帖时间

回复时间

回复数

搜索版块

全部版块

iPhone 4s 综合讨论区(458)

iPhone 4 综合讨论区(443)

iPhone 5s 综合讨论区(394)

iPhone 5 综合讨论区(346)

iPhone 6 综合讨论区(225)

更多 >

搜索时间

全部时间

一天内

一周内

一个月

三个月

半年内

一年内

二年内

三年内

O2O类：淘点点





OSTV类：天猫魔盒



其他

企业内搜索、任何有搜索的场景

名词解释

应用管理

名称	说明
应用	应用是用户的一套数据配置，包括应用的数据源结构，索引结构及其它一些数据属性配置。一个应用即一个搜索服务。
文档	文档是可搜索的结构化数据单元。文档包含一个或

	多个字段，但必须有主键字段，OpenSearch通过主键值来确定唯一的文档。主键重复则文档会被覆盖。
字段	字段是文档的组成单元，包含字段名称和字段内容。
插件	为了方便用户在导入过程中进行一些数据处理，系统内置了若干通用数据处理插件，可以在定义应用结构或者配置数据源的时候通过“内容转换”进行选择。
源数据	用户的原始数据，包含一个或多个源字段。
源字段	组成源数据的最小单元，包含字段名称和字段值，分为文本类型、整型、浮点型三个类型。
索引	索引是一种用于加速文档检索速度的数据结构，一个用户可以创建多个索引。
组合索引	允许用户将多个TEXT、SWS_TEXT等文本类型的源字段索引到同一个字段，用来做组合索引。如一个论坛搜索，需要提供基于标题(title)的搜索及基于标题(title)和内容(body)的综合搜索，那么可以将title建立title_search、default的索引，将body建立default索引。那么，在title_search上查询即可实现基于标题的搜索，在default上查询即可实现基于标题和内容的综合搜索。
索引字段	可以在query子句中使用，需要定义索引字段，通过索引字段来做高性能的检索召回。
属性字段	可以在filter、sort、aggregate、distinct子句使用，用来实现过滤统计等功能。
默认展示字段	用来做结果展示使用，同时可以通过API参数fetch_fields来控制每次结果的返回字段，需注意在程序中配置fetch_fields该参数后会覆盖应用中默认展示字段配置，以程序中的fetch_fields设置为主，若程序中不设置fetch_fields参数则以应用中默认展示字段为主。
分词	对推送上来的文档进行词组切分，TEXT类型为按检索单元进行切分，SWS_TEXT为按单字进行切分。如“浙江大学”，TEXT类型会切分成2个词组：“浙江”、“大学”。SWS_TEXT会切分成4个词组：“浙”、“江”、“大”、“学”。
term	分词后的词组称为term。
构建索引	分完词后会进行索引构建操作，以便根据用户查询，快速定位到具体的文档。搜索引擎一般会构建出两种类型的链表：倒排和正排链表。
倒排	词组到文档的对应关系组成的链表，勾选可搜索后会构建倒排链表。term1->doc1,doc2,doc3；term2->doc1,doc2
正排	文档到字段对应关系组成的链表，勾选可过滤后会构建正排链表。doc1->id,type,create_time...
召回	通过用户查询的关键词进行分词，将分词后的词组

	通过查找倒排链表快速定位到文档，这个过程称为召回。
召回量	召回得到的文档数为召回量。

数据同步

名称	说明
数据源	数据来源，目前系统支持一些主流存储产品的自动对接。
索引重建	重新构建索引数据。一般在首次配置数据源、修改数据源、修改应用结构后需要手动索引重建。定时索引重建一般用于全量数据的重新导入（需要关联数据导入）。

配额管理

名称	说明
文档容量	应用中各个表的总文档大小累加值（不考虑字段名，字段内容按照string来计算容量）。
QPS	每秒查询请求数。

搜索

名称	说明
排序表达式	排序表达式是用于控制搜索结果文档排序的数学表达式，支持基本数学运算、数学函数和内置函数。
粗排表达式	对搜索结果进行第一轮的海选，因为要遍历所有的文档（目前上限为100万），所以粗排要尽量简单（选取对文档最重要的几项内容，如新闻类可以选用文本性及时效性），按照表达式对文档进行算分，并按照算分结果进行排序。
精排表达式	对第一轮的排序结果选取前N个按照精排表达式进行第二轮更细节的分值计算，按照分值进行最终的排序，并返回给用户。
结果摘要	文本内容一般会很长，在搜索结果展示的时候可以只展示部分匹配的内容，方便用户快速了解文档主要内容。
查询分析	可以配置若干分析规则，目前支持拼写检查、停用词、词权重等功能，可以让用户更好的干预搜索行为，获得更好的搜索体验。

发布历史

发布历程

2016-07

- **新版控制台**发布，主要解决相关功能体验问题：

- 1) 通过数据库直接创建应用结构；
- 2) 通过json文件直接生成文档结构；
- 3) 索引字段配置优化设计，方便理解。

2015-10

- 排序表达式新增若干归一化函数，方便用户对不同值域维度的排序因子做权重调整；

2015-09

- 新增**SHORT_TEXT**短文本模糊搜索，支持拼音、前后缀以及单词内模糊搜索效果；

2015-08

- **数据清理**功能发布，支持手动及定期过期数据删除；

- 新增**ENG_TEXT**英文分词及若干匹配相关性feature功能，为英文搜索场景提供更好的搜索效果；

2015-07

- 新增**短信预警**功能，监控文档容量及QPS使用情况，以便用户及时调整配额；

- Java SDK 2.1.3及PHP SDK 2.0.6发布，支持下拉提示、scroll接口、查询分析等接口。

2015-06

- **下拉提示**重磅发布，支持拼音、简写、全拼等模糊搜索功能，为用户提供更加流畅的搜索体验；

- **RDS分表**功能上线，支持简单正则及通配符功能，省时省力并降低误操作概率；

2015-05

- 查询分析增加**同义词**功能，对用户查询词进行同义词改写，提高召回率及搜索结果展现效果；
- **青岛机房上线**，方便青岛区域用户使用；

2015-04

- 为了满足某些场景下对于大批量结果的需求，搜索接口开放了**scroll**功能；

2015-03

- **北京机房上线**，方便北京区域用户使用；
- 新增**多粒度MWS_TEXT及用户自定义NWS_TEXT分词字段类型**，提供更强大的分词功能，提高召回率。
- **Query子句新增Query Boost功能**，允许用户指定查询词权重，**新增exact_match_boost、first_phase_score函数**，二者配合使用，可以更好的控制排序效果。

2015-02

- QP一期重磅发布，新增**提供公共词典的纠错、停用词、词权重等查询分析功能**，可以让用户更好的干预搜索结果，以获得更好的搜索体验；
- **ODPS支持分区等功能**，方便用户做全量数据版本控制；

2015-01

- **杭州机房上线**，提供内网访问域名；
- 错误码优化，方便用户排查问题；
- **RDS数据源自动对接OpenSearch**发布，支持杭州区域mysql5.5、mysql5.6的自动数据同步；
- 应用结构流程优化，数据源及索引重建功能拆分，优化用户导入数据体验；

2014-12

- **附表数据自动更新功能**发布，无需用户手动触发主表即可生效。

2014-11

- **实时引擎发布，数据处理时效性（从数据推送到系统到能检索到）达到10+秒级别**；
- 创建、修改应用结构流程优化：去掉多值字段，改为新增ARRAY类型，更方便用户理解；去掉可聚合类型，目前引擎对于可聚合与可过滤处理逻辑相同，勾选可过滤即可支持原可聚合的功能。

2014-10

- **分区域部署上线**，目前外网仅支持杭州区域，更多区域敬请期待；
- JAVA SDK V2.1.1发布：优化调试信息打印、增加超时时间设置、gzip方式支持、最大连接数设置、开放reranke_size参数等；
- PHP SDK V2.0.4发布：取消默认时区、优化调试信息打印、增加超时时间设置、开放rerank_size参数等；
- 配额预警功能上线，文档即将超配额自动邮件通知，提醒用户及时修正。

2014-09

- **ODPS源自动对接OpenSearch上线**；
- in_polygon、in_query_polygon等新插件function发布，为O2O用户提供更优质的搜索效果体验；
- API/SDK推送文档用户指定时间戳保序功能发布，方便多线程用户更好的控制文档更新流程；

2014-07

- OpenSearch新版控制台上线，展现样式全面改版，全面对接阿里云；
- **OpenSearch上云发布**，开启高性能、可定制搜索托管服务新篇章。