|  |  |  |
| --- | --- | --- |
| 文件状态：  [ ] 草稿  [√] 正式发布  []正在修改 | 文件标识： | SYZH1104-TYPT -11-04 |
| 当前版本： | V1.0 |
| 作者： | 顾洋 |
| 完成日期： | 2018/05/30 |

基于语义网的站内搜索引擎模块改造项目积分系统部署手册

**北京慧博科技有限公司**

**2018年5月**

修订历史版本

|  |  |  |
| --- | --- | --- |
| **日期** | **修订内容** | **修订人** |
| 2018-05-30 | 创建文档 | 顾洋 |
| 2018-05-30 | 审核文档 | 贾世龙 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

目录

第一章 引言 1

1.1 文档目的 1

1.2 预期读者 1

第二章 系统环境要求 2

2.1 硬件列表 2

2.2 软件列表 2

2.3 存储容量规划 2

第三章 系统部署 3

3.1 系统部署方案 3

3.1.1 系统部署计划 3

3.2 系统部署内容 3

3.2.1 搜索部署 3

3.2.2 SSE部署 5

3.2.3 数据同步脚本（ec\_d2r）部署 6

3.3 系统部署验证 11

1. 引言
   1. 文档目的

本文档详细介绍了系统的部署及配置方法。

* 1. 预期读者
* 研发相关人员
* 运维相关人员
* 系统管理员

1. 系统环境要求
   1. 硬件列表

|  |  |  |
| --- | --- | --- |
| 项目 | 最低配置要求 | 推荐配置 |
| 知识图谱服务器 | 4cpu 内存16g 硬盘1T | 8cpu 内存32g 硬盘1T |
| 检索索引服务器 | 4cpu 内存16g 硬盘1T | 8cpu 内存32g 硬盘1T |
| WEB服务器 | 4cpu 内存16g 硬盘1T | 8cpu 内存32g 硬盘1T |
| 日志服务器 | 4cpu 内存16g 硬盘1T | 8cpu 内存32g 硬盘1T |

* 1. 软件列表

本次部署所需服务器软件列表：

|  |  |
| --- | --- |
| **软件名称** | **软件及版本** |
| Tokumx | 2.0.2 |
| Redis | 3.2.8 |
| Tomcat | 8 |
| Nginx | 1.10 |
| Mysql | 5.6 |
| ElasticSearch | 5.5 |
| Logstash | 5.5 |

* 1. 存储容量规划

|  |  |
| --- | --- |
| **部署应用** | **CPU、内存、硬盘大小** |
| 知识图谱服务器 | 8核，内存64G，硬盘：1T |
| 检索索引服务器 | 8核，内存32G，硬盘：1T |
| WEB服务器 | 8核，内存32G，硬盘：1T |
| 日志服务器 | 8核，内存32G，硬盘：1T |

1. 系统部署
   1. 系统部署方案
      1. 系统部署计划

公共数据资源池系统部署的主要工作，包括部署服务器准备、系统运行环境部署、数据库部署、应用安装部署、接口程序部署、接口程序配置调试：

（1）部署服务器准备：收集服务器信息，形成清单，完善至部署手册中，指导部署阶段工作开展；

（2）系统运行环境部署：JDK安装、Tomcat安装；

（3）数据库部署： Mysql数据库安装，Mongodb数据库安装、Solr数据库安装；

（4）应用安装部署：部署应用程序包、配置数据库连接文件、配置应用程序路径；

（5）接口程序部署：完成融合场景接口相关模型及接口部署；

（6）接口程序配置调试：完成接口程序配置，包括执行方式（定时、手动）、源和目标数据源选择、加载方式等。

* 1. 系统部署内容
     1. 搜索部署

（一）部署依赖

JDK1.8

Tomcat7

ElasticSearch 5.4

Mysql

（二）启动步骤

1. 安装并配置jdk1.8环境变量

2. 安装tomcat7

3. 安装并启动elasticsearch5.4，启动命令./bin/elasticsearch &

4. 安装并启动mysql，将esgcc库中的数据恢复

5. 启动tomcat，启动命令./bin/startup.sh

6. 日志查看，/log/catalina.out 或 /project\_logs/

（三）部署配置

1. db.properties

/opt/apache-tomcat-7.0.67/webapps/search/WEB-INF/classes

# elasticsearch ip、集群名、连接端口  
es\_ip=127.0.0.1,127.0.0.1 //重要  
es\_name=my-application  
es\_port=9300  
​# mysql driver、url、user、password  
db.driver=com.mysql.jdbc.Driver  
db.url=jdbc:mysql://127.0.0.1:3306/esgcc?characterEncoding=utf8&amp;autoReconnect=true&amp;failOverReadOnly=false //重要  
db.user=root //重要  
db.password=root //重要  
​

2.meta\_ws.properties

/opt/apache-tomcat-7.0.67/webapps/search/WEB-INF/classes

# 接口文档ip，不需修改

swagger\_ip\_port=127.0.0.1:8080

# elaticsearch 表名、库名、分词器

search\_es\_index=gwds\_prod

search\_es\_type=gwds\_prod\_data

search\_es\_analyzer=ik\_max\_word

# 调用知识图谱接口服务地址url

search\_kg\_semantic\_http\_url=http://127.0.0.1:8080/SSE2/ws/esgcc/semanticSeg //重要改ip port

# 调用知识图谱图名称

search\_kg\_name=u260\_graph\_9ad2fabd //重要,参见本节底部

# 搜索过滤筛选最大返回个数，分品类、商铺、品牌、属性

search\_max\_cat\_num=30

search\_max\_store\_num=30

search\_max\_brand\_num=30

search\_max\_filter\_num=30

# 搜索排序最小分值

sort\_min\_score=10

# 搜索词长boost

weight\_term\_length\_boost=2.0

# 搜索词长boost（命中用户输入）

weight\_segment\_match\_user\_input\_boost=6.0

# 搜索品类boost

weight\_category\_boost= 10.0

# 搜索品类boost（命中用户输入）

weight\_category\_boost\_match\_user\_input=15.0

# 搜索品牌boost

weight\_brand\_boost=8.0

# 搜索销量函数平滑因子

search\_snum\_func\_factor=10

# 搜索评论函数平滑因子

search\_cnum\_func\_factor=10

# 搜索销量函数加权因子

search\_snum\_func\_weight=1

# 搜索评论函数加权因子

search\_cnum\_func\_weight=0.2

# 最大页数及单页最大限制

search\_max\_pageno=50

search\_max\_pagesize=50

# 搜索高亮词前后缀

search\_highlight\_prefix=<em>

search\_highlight\_suffix=</em>

# elasticsearch RESTful地址

es\_url=http://127.0.0.1:9200/ //重要 ip port

# elasticsearch提示词表名（带分站）

prompt\_es\_index=gwds\_prompt\_es\_index

# elasticsearch提示词表名（不带分站）

prompt\_es\_full\_index=gwds\_prompt\_es\_full\_index

# 提示词结果是否排除输入文本

prompt\_exclude\_input=false

# 缺省推荐搜索词列表（unicode编码，半角逗号分隔）

related\_search\_word\_default=\u7535\u8111,\u7A7A\u8C03,\u6253\u5370\u673A,\u7167\u76F8\u673A,\u79FB\u52A8\u786C\u76D8

3.其他配置项

prompt\_query.json 提示词生成规则

related\_search\_query 推荐搜索词生成规则

* + 1. SSE部署

（一）部署依赖

JDK1.8

Tomcat7

Tokumx2.0.2

（二）启动步骤

1. 安装并配置jdk1.8环境变量

2. 安装tomcat7

3. 安装并启动tokumx2.0.2，启动命令./start-tokumx.sh

4. 启动Tomcat，启动命令./bin/startup.sh

（三）部署配置

配置文件路径：SSE2/WEB-INF/lib/kg-service-common-0.1-SNAPSHOT.jar

sse\_config.properties

# 选择知识图谱底层存储介质，并配置相应的参数

# 其中1 mysql; 2 mongodb

kg\_db\_type = 2

kg\_db\_servers = 127.0.0.1 //重要

kg\_db\_port = 19130 //重要

kg\_db\_name = u260\_graph\_9ad2fabd //重要

# 是否启动redis存储，默认不启动

cache\_flag = 0

redis\_servers =

# 实体详细信息读取源

# 0 read from source, 1 read all field in kg

null\_field\_mode = 1

source\_db\_servers = 127.0.0.1 //重要

source\_db\_port = 19130 //重要

source\_db\_name = u260\_graph\_9ad2fabd //重要

# 标注库配置

work\_dir = /work/kg/

tagging\_db\_servers = 127.0.0.1 //重要

tagging\_db\_port = 19130 //重要

tagging\_db\_name = u260\_graph\_9ad2fabd //重要

# swagger配置

swagger\_open = false

swagger\_host = localhost:8080

swagger\_basepath = /SSE2/ws/

* + 1. 数据同步脚本（ec\_d2r）部署

（一）服务器部署环境

服务器：10.1.33.80

项目路径： /opt/es\_d2r

（二）部署依赖

JDK1.8

Python2.7

（三）启动步骤

1. 安装并配置jdk1.8环境变量

2. 安装python2.7

3. 进入es\_d2r目录，运行./centos\_install\_dep.sh安装python第三方依赖库

4. 配置crontab设置定时运行同步脚本sync\_es\_data.sh和sync\_kg\_data.sh

（四）部署配置

1. crontab定时任务

每隔1小时运行同步脚本：

crontab -e

# 添加以下两条命令(注意修改ec\_d2r目录)

0 \* \* \* \* cd {ec\_d2r目录}; sh sync\_es\_data.sh >> /var/log/sync\_es\_data.log 2>&1

0 \* \* \* \* cd {ec\_d2r目录}; sh sync\_kg\_data.sh >> /var/log/sync\_kg\_data.log 2>&1

2.知识图谱配置

编辑ec\_d2r/config.json(配置内容如下)，根据实际环境配置oracle、mongodb和mysql的连接配置。

{

# oracle商品源数据库配置

"oracle": {

"host": "127.0.0.1",

"port": 1521,

"sid": "XE",

"user": "ESGCC\_DBUSER",

"password": "123456"

},

# 知识图谱mongo存储配置

"mongodb": {

"host": "127.0.0.1",

"port": 27017

},

# 商品品牌、分类、属性映射表

"mysql": {

"host": "120.27.14.254",

"port": 3306,

"user": "root",

"password": "root@hiekn",

"db": "esgcc"

}

}

3. 商品搜索提示配置

编辑ec\_d2r/es\_data\_export/prompt/esgcc-prompt.properties(配置内容如下)，修改搜索提示引擎es配置。

es\_ip=120.27.14.254

es\_cluster\_name=my-application

prompt\_es\_index=gwds\_prod\_prompt\_test

prompt\_es\_full\_index=gwds\_prod\_prompt\_full\_test

category\_attribute\_file\_path=category\_attribute\_ori.txt

category\_syn\_file\_path=category\_syn.txt

brand\_syn\_file\_path=brand\_syn.txt

prompt\_es\_setting\_path=gwds\_prod\_prompt\_test\_settings.json

prompt\_es\_mapping\_path=gwds\_prod\_prompt\_test\_mapping.json

input\_file\_path=../datas/

output\_file\_path=output/

4. 商品搜索配置

编辑ec\_d2r/es\_data\_export/search/config.json(配置内容如下)，根据实际搜索服务器修改es和mysql连接配置项。

{

# 商品索引引擎es

"es": {

"es\_name": "my-application",

"es\_ip": "120.27.14.254",

"es\_port": 9300,

"index\_name": "gwds\_prod\_a",

"type\_name": "gwds\_prod\_v1\_data",

"es\_insert\_batch\_size": 2000,

"es\_update\_batch\_size": 100,

"prod\_index\_max\_num": 0

},

# 商品品牌、分类、属性映射表

"mysql": {

"db\_url": "jdbc:mysql://120.27.14.254:3306/esgcc?characterEncoding=utf8&autoReconnect=true&failOverReadOnly=false&useSSL=false",

"db\_user": "root",

"db\_password": "root@hiekn"

},

"es\_index\_setting": {

"number\_of\_shards": 1,

"number\_of\_replicas": 0,

"analysis": {

"analyzer": {

"by\_max\_word": {

"filter": [

"lowercase",

"unique"

],

"type": "custom",

"tokenizer": "ik\_max\_word"

}

}

},

"similarity": {

"my\_similarity": {

"type": "BM25",

"k1": "0"

}

}

},

"es\_type\_mapping": {

"properties": {

"prod\_id": {

"store": true,

"type": "keyword"

},

"prod\_name\_toshow": {

"store": true,

"type": "keyword"

},

"prod\_adword": {

"store": true,

"type": "keyword"

},

"prod\_name": {

"analyzer": "by\_max\_word",

"similarity": "my\_similarity",

"store": true,

"type": "text"

},

"prod\_division": {

"type": "nested",

"properties": {

"name": {

"type": "keyword"

},

"id": {

"type": "keyword"

}

}

},

"prod\_site": {

"type": "nested",

"properties": {

"name": {

"type": "keyword"

},

"id": {

"type": "keyword"

}

}

},

"prod\_has\_site": {

"store": true,

"type": "keyword"

},

"prod\_price": {

"store": true,

"type": "double"

},

"prod\_origin\_price": {

"store": true,

"type": "double"

},

"prod\_putaway\_time": {

"store": true,

"type": "date",

"format": "yyyy-MM-dd HH:mm:ss"

},

"prod\_status": {

"store": true,

"type": "keyword"

},

"prod\_brand\_name": {

"store": true,

"type": "keyword"

},

"prod\_store\_id": {

"store": true,

"type": "keyword"

},

"prod\_store\_name": {

"store": true,

"type": "keyword"

},

"prod\_cat\_name": {

"store": true,

"type": "keyword"

},

"prod\_cat\_name\_lv2": {

"store": true,

"type": "keyword"

},

"prod\_cat\_name\_lv1": {

"store": true,

"type": "keyword"

},

"snum": {

"store": true,

"type": "long"

},

"cnum": {

"store": true,

"type": "long"

},

"prod\_img\_url": {

"store": true,

"type": "keyword"

},

"prod\_material\_id": {

"store": true,

"type": "keyword"

},

"prod\_props": {

"type": "nested"

}

}

}

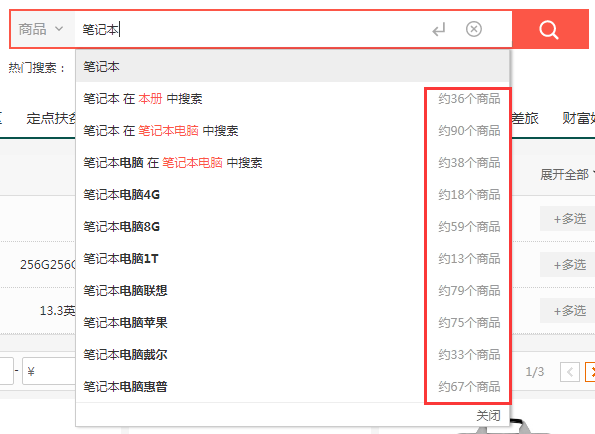
}

* 1. 系统部署验证

系统部署完成后，访问系统页面查看是否能够正常访问，正常搜索笔记本、空调等是否有结果。能够正常访问且能进行正常搜搜，即表示验证通过。

测试案例1：

在搜索框输入笔记本，下拉提示框返回搜索词，提示搜索词所属范围，统计并返回商品数量。如下图所示：



测试案例2：

在搜索框输入奥克斯空调，搜索返回结果为空调，品牌为奥克斯。

