

ALIBABA CLOUD

阿里云

图数据库
全文检索

文档版本：20220704

 阿里云

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
 危险	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 危险 重置操作将丢失用户配置数据。
 警告	该类警示信息可能会导致系统重大变更甚至故障，或者导致人身伤害等结果。	 警告 重启操作将导致业务中断，恢复业务时间约十分钟。
 注意	用于警示信息、补充说明等，是用户必须了解的内容。	 注意 权重设置为0，该服务器不会再接受新请求。
 说明	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 说明 您也可以通过按Ctrl+A选中全部文件。
>	多级菜单递进。	单击设置> 网络> 设置网络类型。
粗体	表示按键、菜单、页面名称等UI元素。	在结果确认页面，单击确定。
Courier字体	命令或代码。	执行 <code>cd /d C:/window</code> 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	<code>bae log list --instanceid</code> <code>Instance_ID</code>
[] 或者 [a b]	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
{ } 或者 {a b}	表示必选项，至多选择一个。	<code>switch {active stand}</code>

目录

1. 简要介绍	05
2. 准备工作	07
3. GDB中配置ES信息	10
4. GDB+ES的DSL编写范例	13
5. ES操作状态查询表	16

1. 简要介绍

图数据库GDB深度融合Elasticsearch(ES)，使得GDB具备了分词检索、模糊匹配、正则表达式查询的能力，且GDB内核集成了数据增量同步至ES中的服务，保证GDB与ES数据的强一致性。本文介绍使用GDB+ES的全文索引服务的方法。

背景介绍

如果您想使用GDB+ES的全文索引服务，建议您购买新的GDB实例，因为目前GDB数据仅支持增量同步至ES中。

当您已经在使用GDB实例，也想使用GDB+ES的全文索引服务，您需要进行以下步骤：

1. 将您的GDB实例升级到最新的版本。
2. 需要将您GDB中数据删除。
3. 购买ES服务并在GDB中注册ES信息。
4. 重新导入数据。

当然，您如果只想对部分数据进行全文索引，那么您也可以只删除部分数据，然后重新导入需要全文索引的数据即可。但是，需要注意的是，GDB数据增量同步至ES对这种形式的支持还不是很友好，建议您购买新的实例。

GDB全文搜索的数据模型：

点模型

```
{
  "id": "Vertex Id",
  "label": [List of Labels type],
  "type": "vertex/edge"
  "predicates": {
    "Property name": [
      {
        "value": "Property Value1",
      },
      {
        "value": "Property Value2"
      }
    ]
  }
}
```

- `id`：文档的实体唯一ID。
- `label`：点或边的一个或多个标签。
- `type`：用于指定当前文档表示点或边。
- `predicates`：对于Gremlin，存储点或边的属性键值，比如name字段在ES中存储为 `predicates.name.value` 的形式。
- `value`：存储属性值，支持GDB中的set属性。

边模型

```
{
  "id": "Vertex Id/Edge Id",
  "label": [List of Labels type],
  "type": "vertex/edge"
  "toId": "入边id",
  "toLabel": "入边label",
  "fromId": "出边id",
  "fromLabel": "出边label",
  "predicates": {
    "Property name": [
      {
        "value": "Property Value",
      },
      {
        "value": "Property Value2",
      }
    ]
  }
}
```

GDB+ES使用流程

1. 查看[前提条件](#)，购买GDB与ES实例。
2. 查看[GDB中配置ES信息](#)，将ES实例信息注入GDB。
3. 查看[GDB+ES的DSL编写范例](#)，选择合适的查询方式进行查询。

2.准备工作

本文介绍使用全文检索功能的前提条件，以及配置基本参数等准备工作。

前提条件

- 已创建阿里云ElasticSearch实例，创建方法请参见[创建阿里云Elasticsearch实例](#)。

说明

- 购买版本在7.4及以上，建议优先选择7.4版本。
- 在设置账号密码时，不要带有&字符。
- 选择磁盘容量大小请参见[容量大小指南](#)。

- 已创建图数据库GDB实例，创建方法请参见[创建主实例](#)。
- 图数据库GDB实例和ElasticSearch实例所属地域和可用区相同。
- （可选）创建与图数据库GDB实例、ElasticSearch实例所属地域和可用区相同，网络类型相同且使用同一个网络的ECS实例，具体请参见[创建ECS实例](#)。

配置基本参数

1. 开启ElasticSearch实例自动创建索引

关闭自动创建索引将导致图数据库GDB的数据无法正常同步至ElasticSearch实例中。

- 登录[阿里云Elasticsearch控制台](#)。
- 在左侧导航栏，单击Elasticsearch实例。
- 在页面左上角，选择ElasticSearch实例所属资源组和地域。
- 单击目标ElasticSearch实例ID或目标ElasticSearch实例所在行操作列的管理。
- 在左侧导航栏，单击配置与管理 > ES集群配置。
- 单击YAML文件配置区域右侧的修改配置。
- 在YAML文件配置面板，将自动创建索引设置为允许自动创建索引。
其他配置请参见[修改配置](#)。

viii. 单击确定。

2. 修改ElasticSearch实例中场景化配置模板

- 登录[阿里云Elasticsearch控制台](#)。
- 在左侧导航栏，单击Elasticsearch实例。
- 在页面左上角，选择ElasticSearch实例所属资源组和地域。
- 单击目标ElasticSearch实例ID或目标ElasticSearch实例所在行操作列的管理。
- 在左侧导航栏，单击配置与管理 > ES集群配置。
- 单击场景化配置区域的索引模板配置（场景化配置模板为通用场景）。

vii. 在索引模板配置面板的当前配置输入框中，执行如下操作。

- 在 settings 后添加以下参数：

索引模板配置	
当前配置	场景化推荐模板（通用场景）
<pre>1 { 2 "order": -2147483647, 3 "index_patterns": [4 "*" 5], 6 "settings": { 7 "number_of_shards": "3" 8 "number_of_replicas": "1" 9 "index": { 10 "search": { 11 "slowlog": { 12 "level": "info",</pre>	<pre>1 { 2 "order": -2147483647, 3 "index_patterns": [4 "*" 5], 6 "settings": { 7 "index": { 8 "search": { 9 "slowlog": { 10 "level": "info", 11 "threshold": { 12 "fetch": {</pre>

说明 如果当前配置为空，请先单击一键应用模板，然后在 settings 后添加参数。

```
"number_of_shards": "3" //ElasticSearch实例的数据节点数
"number_of_replicas": "1"
```

说明 您可以在ElasticSearch实例的基本信息页面查看数据节点数。

- 将 refresh_interval 设置为1s。

索引模板配置	
当前配置	场景化推荐模板（通用场景）
<pre>18 "info": "100ms" 19 }, 20 "query": { 21 "warn": "500ms", 22 "trace": "50ms", 23 "debug": "100ms", 24 "info": "200ms" 25 } 26 } 27 } 28 }, 29 "refresh_interval": "1s", 30 "unassigned": { 31 "node_left": { 32 "delayed_timeout": "5m"</pre>	<pre>16 "info": "100ms" 17 }, 18 "query": { 19 "warn": "500ms", 20 "trace": "50ms", 21 "debug": "100ms", 22 "info": "200ms" 23 } 24 } 25 } 26 }, 27 "refresh_interval": "10s", 28 "unassigned": { 29 "node_left": { 30 "delayed_timeout": "5m"</pre>

其他配置请参见索引模板配置。

viii. 单击提交。

3. （可选）选择以下任意一种方法配置ElasticSearch实例基本参数（能够有效的防止触发ElasticSearch熔断机制）。
 - 发送curl请求配置参数
 - a. 登录ECS实例，具体请参见连接方式概述ECS远程连接操作指南。

b. 执行如下命令，发送curl配置参数。

```
curl -u elastic:${your-es-password} -XPUT "http://${your-es-endpoint}:9200/_cluster/settings" -H 'Content-Type: application/json' -d '{
  "persistent" : {
    "indices.breaker fielddata.limit" : "40%",
    "indices.breaker.request.limit" : "40%",
    "indices.breaker.total.limit" : "70%"
  }
}'
```

② 说明

- `${your-es-endpoint}`：请替换为ElasticSearch实例的域名。
- `${your-es-password}`：请替换为ElasticSearch实例的密码。

o 在Kibana控制台配置参数

a. 登录Kibana控制台，具体请参见[登录Kibana控制台](#)。

b. 执行如下命令，配置参数。

```
PUT /_cluster/settings {
  "persistent" : {
    "indices.breaker fielddata.limit" : "40%",
    "indices.breaker.request.limit" : "40%",
    "indices.breaker.total.limit" : "70%"
  }
}
```

3.GDB中配置ES信息

购买GDB与ES实例后，需要在GDB实例中注册ES信息。

1. 登录ECS实例，具体请参见[连接方式概述ECS远程连接操作指南](#)。

2. 注册并查看ES信息。

i. 执行如下命令，在GDB中注册ES信息。

```
curl -u ${username}:${password}
"http://${your-gdb-endpoint}:8182/controller/es/updateEsService?endpoint=${your-es-
endpoint}&port=${your-es-port}&username=${your-es-username}&password=${your-es-pass
word}"
```

- 将 `${username}` 改为您的图数据库GDB实例的用户名。
- 将 `${password}` 改为您的图数据库GDB实例的密码。
- 将 `${your-gdb-endpoint}` 改为您的图数据库GDB实例的域名，如果您用外网访问的方式，下面的8182端口要替换为3734。
- 将 `${your-es-endpoint}` 替换为您的ES实例的域名。
- 将 `${your-es-port}` 改为您的ElasticSearch实例的端口，默认端口应为9200。
- 将 `${your-es-username}` 改为你的ElasticSearch实例的用户名，默认名应为elastic。
- 将 `${your-es-password}` 改为您的ElasticSearch实例的密码。
- 您在指定每个字段的时候，不要传入空格，否则语句将无法执行，且无任何返回值。

ii. 执行如下命令，查看注册ES信息。

您的ES账号的密码已经被加密存储。

```
curl -u ${username}:${password} "http://${your-gdb-endpoint}:8182/controller/es/get
EsService"
```

iii. 查看与ES连接的状态。

在导入数据之前，一定要执行这一步，这个主要是通过connectStatus字段观察配置的信息，具体返回的信息，以及对对应状态的含义可以参考附表。

```
curl -u ${username}:${password}
"http://${your-gdb-endpoint}:8182/controller/es/getMonitor"
```

3. 注册并查看同步字段信息。

o 指定同步字段。

在你第一次使用ES服务的时候，需要指定需要增量同步至ES的属性字段，这样可以减轻同步链路的负担。

- 将 `vertex-property` 替换为您数据集中点的属性字段。
- 将 `edge-propoerty` 替换为您数据集中边的属性字段。
- 指定的字段中不能包含 “,” 号和 “.” 点，否则无法指定成功。
- 不需要指定label的字段，其已经默认同步至ES中。

- 如果您刚开始并不想指定字段，也需要运行该语句，要不然ES服务将不可用，您只需要在vertex与edge中不指定同步字段就好，即 `"vertex":[""],"edge":[""]` 这样的形式（双引号内不要有空格），将会同步点边属性所有的字段至ES中,但是包含 “,” 号和 “.” 点的属性不会同步。
- 只支持字符串类型的属性同步至ES中，如果您的属性的字段是非字符串类型，例如，整形、浮点型等，这些字段将不会同步至ES中，即使您指定了需要将该字段同步至ES中。
- 您如果在使用GDB的服务中，想指定一个新的字段同步至ES中，您可以重新执行该操作，加入需要指定的字段即可，但是需要特别注意的是，本操作是覆盖操作，因此您需要将以前指定的同步字段也一并写入，还要注意该操作一定要在向GDB写入数据之前，因为这样可以确保您新添加的字段能够正确的同步至ES中。

```
curl -u ${username}:${password} -H "Content-Type: application/json" -X POST --data '{
  "vertex":["${vertex-property1}","${vertex-property2}"],
  "edge":["${edge-property1}","${edge-property2}"]
}' "http://${your-gdb-endpoint}:8182/controller/es/setField"
```

○ 查看指定同步字段。

可以查看点边指定的导入字段，如果发现指定字段错误，可以重新指定需要导入的字段。

```
curl -u ${username}:${password} "http://${your-gdb-endpoint}:8182/controller/es/getField"
```

返回结果如下：

- `vertex` 表示点属性同步至ES中的字段。
- `edge` 表示边属性同步至ES中的字段。

```
{"payload":{"edge":{"weight":"","desc":"","vertex":{"value":"","name":"","age":"","location":""}}},"status":"200 OK"}
```

4. 开启同步，将GDB数据增量同步至ES中。

如果您配置完ES的信息，并指定完点边需要导入的字段后，您就可以通过以下命令将GDB的数据增量同步至ES中了。

```
curl -u ${username}:${password} "http://${your-gdb-endpoint}:8182/controller/es/beginSync"
```

5. 查看点和边在ES中的index取值。

该步骤不是必要的步骤，如果您只是想使用GDB+ES,不关注数据在ES中的存储，这一步骤可以跳过。但您想在ES中查询数据同步的情况，这一步又是必须的。

```
curl -u ${username}:${password} "http://${your-gdb-endpoint}:8182/controller/es/getEsIndex"
```

返回结果如下：

- `gdbVertex` 表示点数据在ES中index的取值。
- `gdbEdge` 表示边数据在ES中index的取值。

```
{"payload":{"gdbEdge":{"gdb_edge_cvkbdqcl"},"gdbVertex":{"gdb_vertex_cvkbdqcl"}},
"status":"200 OK"}
```

6. 查看同步链路的状态。

通过以下命令，可以查看GDB同步至ES中数据链路的状态。

```
curl -u ${username}:${password} "http://${your-gdb-endpoint}:8182/controller/es/getSyncStatus"
```

4.GDB+ES的DSL编写范例

本文主要为您提供使用图数据库GDB+ES（Elasticsearch，搜索引擎）的DSL编写指南。

使用建议

- GDB#match与GDB#phrase相比较，GDB#match查询范围广，一般情况下，GDB#match搜索结果的前几名即GDB#phrase的搜索结果。所以，GDB#match的模糊匹配能力大于GDB#phrase。

② 说明

- GDB#match：注重查询的广度。
- GDB#phrase：注重查询的精度。

- 如果您有 `contain` 查询需求，建议您先使用GDB#phrase（精度应该能满足您绝大部分的查询需求），如果您对搜索结果不满意，您可以使用GDB#regexp正则表达式进行查询（耗时长）。

基本查询形式

```
g.withSideEffect('GDB#fts','enable').V().has('${property-name}', 'GDB#* ${your-search-value}').limit(topN)
```

- `g.withSideEffect('GDB#fts','enable')` 表示DSL需要使用GDB+ES的全文检索。
- `V()` 表示查询的为顶点的属性。如果是 `E()`，表示查询为边的属性。
- `has()` 为GDB+ES的必要步骤。

② 说明

- value以 `'GDB#*'` 开头，其中请将*替换为 `match、phrase` 等。
- `property-name`：请替换为您需要搜索的字段。
- `your-search-value`：请替换为您需要搜索内容。其中，`'GDB#*'` 与 `${your-search-value}` 之间必须有一个空格。

- `limit(topN)`：可以控制返回的条数，默认按照打分结果排序返回10条。即使是部分走ES查询，`limit()`所限定的返回数也是经由ES查询的返回条数。因此，您的查询需求如果是先通过 `GDB+ES` 的全文检索功能检索出特定的点，再对该点执行其他操作，需要适当调大 `limit()`。
- 可以加上 `hasLabel()` 限定。

GDB#match

使用了ES中的match查询，为分词查询、模糊匹配、结果会根据相关度打分，搜索范围比较广，使得GDB具备了分词查询的能力，基本的查询形式为：

```
g.withSideEffect('GDB#fts','enable').V().has('${property-name}', 'GDB#match ${your-search-value}').limit(topN)
```

示例：查询desc字段中包含优秀人才的点

```
g.withSideEffect('GDB#fts','enable').V().has('desc','GDB#match 优秀人才').limit(2)
==>[desc:[新世纪优秀人才支持计划]]
==>[desc:[全国优秀艺术人才]]
```

GDB#phrase

使用了ES中的 `match_phrase` 查询，也会进行分词查询，搜索范围比较小，搜索结果更加精准，类似GDB中的 `TextP.containing()`。

```
g.withSideEffect('GDB#fts','enable').V().has('${property-name}', 'GDB#phrase ${your-search-value}').limit(topN)
```

示例：查询desc字段中包含优秀人才的点

```
g.withSideEffect('GDB#fts','enable').V().has('desc','GDB#phrase 优秀人才').limit(2)
==>[desc:[新世纪优秀人才支持计划]]
```

GDB#regexp

❓ 说明 在实际生产中，不建议您使用这种搜索形式。因为该搜索形式需要全库扫描索引，非常耗时。如果您有这种包含查询的需求，建议您使用GDB#phrase查询，其查询的效果应该能满足您的大部分需求。当您发现使用GDB#phrase与GDB#match都无法满足您的查询需求时使用该搜索形式的包含查询。

使用了ES中的 `regexp` 正则查询，使得GDB具备了正则查询的能力，`regexp`正则查询功能非常强大。

```
g.withSideEffect('GDB#fts','enable').V().has('${property-name}', 'GDB#regexp ${your-search-value}').limit(20)
```

示例1：查询location字段中以深圳开头的点

```
g.withSideEffect('GDB#fts','enable').V().has('location', 'GDB#regexp 深圳.*').limit(topN)
```

示例2：查询location字段中以深圳结束的点

```
g.withSideEffect('GDB#fts','enable').V().has('location', 'GDB#regexp .*深圳').limit(topN)
```

示例3：查询location字段包含深圳的点

```
g.withSideEffect('GDB#fts','enable').V().has('location', 'GDB#regexp .*深圳.*').limit(topN)
```

GDB#string

❓ 说明 GDB#string为分词查询。

该查询使用了ES中的 `query_string`，支持 `AND OR NOT` 等操作，详细信息请参见[query_string链接](#)。

```
g.withSideEffect('GDB#fts','enable').V().has('${property-name}', 'GDB#string ${your-search-value}').limit(topN)
```

示例：查询 `location` 字段是深圳或者武汉的点

```
g.withSideEffect('GDB#fts','enable').V().has('location', 'GDB#string 深圳 OR 武汉').limit(topN)
```

GDB#expression

② 说明

- GDB中 `property` 的字段，在ES中的存储模型是不一样的，例如，`name` 在ES中存储为 `predictates.name.value` 的形式。
- 加上 `.keyword` 为等值查询；不加为分词查询，效果同GDB#match。

GDB#expression 使用的也是 `query_string`，查询形式如下：

```
g.withSideEffect('GDB#fts','enable').V().has('*', 'GDB#expression ${your-search-expression}').limit(topN)
```

② 说明

- 其与 `GDB#string` 的区别是没有指定查询字段，用*代替。
- `${your-serach-expression}`：请替换为具体的查询表达式。

GDB#nokey

GDB#nokey的查询效果和GDB#phrase类似，区别是GDB#nokey不需要指定搜索主键。

```
g.withSideEffect('GDB#fts', "enable").V().has('*', 'GDB#nokey ${your-search-value}').limit(topN)
```

5.ES操作状态查询表

本文档主要介绍您在使用GDB+ES时，执行各命令的返回值以及对应的操作指南。

操作	返回值	操作指南
查询ES状态，或开启同步任务时	"connectStatus":"no es configuration information"	说明没有配置ES客户端的信息，需要及时配置客户端信息。
查询ES状态时或开启同步任务时	"connectStatus":"green","info":"client connection normal, all services are accessible"	说明当前与ES连接状态良好，所有的ES分片都可用。
查询ES状态，或开启同步任务时	"connectStatus":"yellow","info":"client connection normal, all services are accessible"	当前与ES是可以连接的，但是ES有部分副本分片服务故障，阿里云ES团队会进行监控并处理。
查询ES状态，或开启同步任务时	"connectStatus":"yellow","info":"client connection normal, all services are accessible"	当前与ES时可以连接的，但是ES有部分主分片服务故障，阿里云ES团队会进行监控并处理。
查询ES状态，或开启同步任务时	"connectStatus":"failure","info":"client connection failure, please check the es configuration information"	GDB服务与ES连接失败，需要仔细检查ES的配置信息，是否存在错误。需要确保GDB服务与ES服务在同一region下的同一可用区。检查完成后，重新配置，如果还是返回该错误，请及时联系技术人员。
开启同步任务时	"connectStatus":"the read binlog and write es threads have run successfully"	增量同步的线程已经启动。
开启同步任务时	"connectStatus":"the read binlog and write es thread are already running"	增量同步的线程正在运行，无需重复执行启动操作。
使用GDB+ES查询时错误	enable misspelling,cannot start full text search service	使用 <code>g.withSideEffect('GDB#fts','enable')</code> 时，enable拼写错误。因此无法正确启动GDB全文检索服务
使用GDB+ES查询时错误	please use the correct GDB#beginning	GDB的全文索引，现在支持GDB#match、GDB#phrase、GDB#regexp、GDB#string、GDB#expression、GDB#nokey等，如果您有新的查询需求，可以联系技术人员。
使用GDB+ES查询时错误	This dsl is illegal, full text by using elasticsearch should be global	使用GDB+ES的DSL针对的是全文索引，你可能错误的使用了该操作、比如加了out()、both()等操作。

操作	返回值	操作指南
使用GDB+ES查询时错误	使用GDB+ES查询时错误	比如在GDB+ES全文索引使用了or(has(),has())这样的形式，GDB+ES全文检索功能暂不支持这样的查询形式。
使用GDB+ES查询时错误	使用GDB+ES查询时错误	在使用GDB+ES全文索引时，指定的属性key值存在暂不支持的字符，比如","号和"."号。