RAG实践（端到端）

思路：以 **workspace 过滤** → **/chunks 语义检索（OpenAI Embedding）** → **拼接上下文** → **调用 LLM** 为主线

0 健康检查

```
curl -s -H "x-api-key: $KEY" "http://127.0.0.1:8000/db" | jq
```

```
(base) [root@ecs-dbd1 ~]# curl -s -H "x-api-key: $KEY" "http://127.0.0.1:8000/db" | jq
"Connected to openGauss (pg)."
```

1 创建工作区

```
cat <<JSON | curl -s -X POST \
  -H "x-api-key: $KEY" -H "Content-Type: application/json" \
  --data-binary @- "http://127.0.0.1:8000/workspaces" | jq
{
  "name": "PG-demo",
  "description": "demo ws"
}
JSON
```

```
(base) [root@ecs-dbd1 ~]# cat <<JSON | curl -s -X POST \
>   -H "x-api-key: $KEY" -H "Content-Type: application/json" \
>   --data-binary @- "http://127.0.0.1:8000/workspaces" | jq
> {
>   "name": "PG-demo",
>   "description": "demo ws"
> }
> JSON
{
  "message": "created",
  "status": "success",
  "workspace": {
    "id": "83a48b27-7ed9-4111-ad19-ca910e477653",
    "name": "PG-demo",
    "created_by": "12345678-1234-1234-1234-1234567890ab",
    "created_at": "2025-09-09T12:31:52.906023+00:00",
    "updated_at": "2025-09-09T12:31:52.906023+00:00"
  }
}
```

```
export WS_ID="83a48b27-7ed9-4111-ad19-ca910e477653"
```

思路：以 **workspace 过滤** → **/chunks 语义检索（OpenAI Embedding）** → **拼接上下文** → **调用 LLM** 为主线

2 创建chunks

```
# 1) 组织 payload
cat >/tmp/chunks_payload.json <<'JSON'
{
  "chunks_in": [
    {
      "content": "this is a text chunk",
      "tags": ["a", "b"],
      "user_metadata": { "score": 0.9 }
    },
    {
      "content": { "k1": "v1", "k2": 2 },
      "tags": ["x"],
      "user_metadata": { "note": "obj" }
    }
  ]
}
JSON

# 2) 创建（按 workspace 归属；只需 x-api-key，不用传 created_by）
curl -s -X POST \
  -H "x-api-key: $KEY" \
  -H "Content-Type: application/json" \
  --data-binary @/tmp/chunks_payload.json \
  "http://127.0.0.1:8000/chunks?workspace_id=$WS_ID" \
| tee /tmp/chunks_create.json | jq

export CHUNK_A=$(jq -r '.chunks[0].id // .chunks_created[0].id'
/tmp/chunks_create.json)
export CHUNK_B=$(jq -r '.chunks[1].id // .chunks_created[1].id'
/tmp/chunks_create.json)
echo "CHUNK_A=$CHUNK_A"
echo "CHUNK_B=$CHUNK_B"

# 3)增加中文chunk

cat >/tmp/chunks_more.json <<'JSON'
{
  "chunks_in": [
    {
      "content": "openGauss 是企业级开源数据库，兼容 PostgreSQL，具备高可用与高性能。",
      "tags": ["db","opengauss"],
      "user_metadata": { "lang": "zh" }
    },
    {
      "content": "WhyHow 支持文档分块、图谱抽取与 RAG 检索问答，便于企业知识应用。",
      "tags": ["whyhow","rag"],
      "user_metadata": { "lang": "zh" }
    }
  ]
}
JSON

curl -s -X POST \
  -H "x-api-key: $KEY" \
  -H "Content-Type: application/json" \
```

```
  --data-binary @/tmp/chunks_more.json \
  "http://127.0.0.1:8000/chunks?workspace_id=$WS_ID" | jq
```

```
(base) [root@ecs-dbd1 ~]# curl -s -X POST \
>   -H "x-api-key: $KEY" \
>   -H "Content-Type: application/json" \
>   --data-binary @/tmp/chunks_payload.json \
>   "http://127.0.0.1:8000/chunks?workspace_id=$WS_ID" \
> | tee /tmp/chunks_create.json | jq
{
  "message": "created",
  "status": "success",
  "count": 2,
  "chunks": [
    {
      "id": "73ee3834-b1b2-4aa1-a4bb-725d03fbeedf",
      "created_at": "2025-09-09T12:53:32.083061Z",
      "updated_at": "2025-09-09T12:53:32.083061Z",
      "created_by": "12345678-1234-1234-1234-1234567890ab",
      "workspaces": [
        "83a48b27-7ed9-4111-ad19-ca910e477653"
      ],
      "document": null,
      "data_type": "string",
      "content": "this is a text chunk",
      "embedding": null,
      "metadata": {
        "language": "en",
        "length": 20,
        "size": 69,
        "data_source_type": "manual",
        "index": null,
        "page": null,
        "start": null,
        "end": null
      },
      "tags": {
        "83a48b27-7ed9-4111-ad19-ca910e477653": [
          "a",
          "b"
        ]
      },
      "user_metadata": {
        "83a48b27-7ed9-4111-ad19-ca910e477653": {
          "score": 0.9
        }
      }
```

检查embedding

```
curl -s -H "x-api-key: $KEY" \
  "http://127.0.0.1:8000/chunks?
workspace_id=$WS_ID&limit=50&populate=true&include_embeddings=true" \
| jq '.chunks[] | {id, type: .data_type, dim: (.embedding|length? // 0), preview:
((.content|tostring)[0:50])}'
```

```
(base) [root@ecs-dbd1 ~]# curl -s -H "x-api-key: $KEY" \
>   "http://127.0.0.1:8000/chunks?workspace_id=$WS_ID&limit=50&populate=true&include_embeddings=true" \
> | jq '.chunks[] | {id, type: .data_type, dim: (.embedding|length? // 0), preview: ((.content|tostring)[0:50])}'
{
  "id": "ce060b91-1bbb-4451-83c8-8a7c52263e5d",
  "type": "string",
  "dim": 1536,
  "preview": "inline content\nline2"
}
{
  "id": "54a88fa4-5124-41ba-addd-84f3d95c33de",
  "type": "string",
  "dim": 1536,
  "preview": "Hello\nThis is a txt chunk."
}
{
  "id": "a3ff672b-4b9e-40bc-b4c1-3b04af941be5",
  "type": "object",
  "dim": 1536,
  "preview": "{\"k1\":\"v1\",\"k2\":2}"
}
{
  "id": "069cf8e0-b80d-4dc1-ba52-4caefd9c7a22",
  "type": "object",
  "dim": 1536,
  "preview": "{\"id\":2,\"name\":\"Bob\",\"score\":0.8}"
}
{
  "id": "69e9f6d8-9791-45a1-9522-983b01a7eb7e",
  "type": "object",
  "dim": 1536,
  "preview": "{\"id\":1,\"name\":\"Alice\",\"score\":0.9}"
}
{
  "id": "8f60aa04-9a9d-4a61-ab17-bd0dcf2e801a",
  "type": "object",
  "dim": 1536,
  "preview": "{\"id\":1,\"name\":\"alpha\",\"score\":0.91}"
}
{
  "id": "d12d782c-0b2e-4813-87a1-8cd26524b999",
  "type": "object",
  "dim": 1536,
  "preview": "{\"id\":2,\"name\":\"beta\",\"score\":0.83}"
}
(base) [root@ecs-dbd1 ~]#
```

```
curl -s -H "x-api-key: $KEY" \
  "http://127.0.0.1:8000/chunks?workspace_id=$WS_ID&limit=50&populate=true" \
| jq '.chunks[] | {id, has_emb: (has("embedding") and .embedding!=null), type:
.data_type, preview: ((.content|tostring)[0:50])}'
```

```
(base) [root@ecs-dbd1 ~]# curl -s -H "x-api-key: $KEY" \
>     "http://127.0.0.1:8000/chunks?workspace_id=$WS_ID&limit=50&populate=true" \
> | jq '.chunks[] | {id, has_emb: (has("embedding") and .embedding!=null), type: .data_type, preview: ((.content|tostring)
[0:50])}'
{
  "id": "ce060b91-1bbb-4451-83c8-8a7c52263e5d",
  "has_emb": true,
  "type": "string",
  "preview": "inline content\nline2"
}
{
  "id": "54a88fa4-5124-41ba-addd-84f3d95c33de",
  "has_emb": true,
  "type": "string",
  "preview": "Hello\nThis is a txt chunk."
}
{
  "id": "a3ff672b-4b9e-40bc-b4c1-3b04af941be5",
  "has_emb": true,
  "type": "object",
  "preview": "{\"k1\":\"v1\",\"k2\":2}"
}
{
  "id": "069cf8e0-b80d-4dc1-ba52-4caefd9c7a22",
  "has_emb": true,
  "type": "object",
  "preview": "{\"id\":2,\"name\":\"Bob\",\"score\":0.8}"
}
{
  "id": "69e9f6d8-9791-45a1-9522-983b01a7eb7e",
  "has_emb": true,
  "type": "object",
  "preview": "{\"id\":1,\"name\":\"Alice\",\"score\":0.9}"
}
{
  "id": "8f60aa04-9a9d-4a61-ab17-bd0dcf2e801a",
  "has_emb": true,
  "type": "object",
  "preview": "{\"id\":1,\"name\":\"alpha\",\"score\":0.91}"
}
{
  "id": "d12d782c-0b2e-4813-87a1-8cd26524b999",
  "has_emb": true,
  "type": "object",
  "preview": "{\"id\":2,\"name\":\"beta\",\"score\":0.83}"
}
```

3 RAG查询

1. GET

```
curl -s -G -H "x-api-key: $KEY" \
  --data-urlencode "workspace_id=$WS_ID" \
  --data-urlencode "text=openGauss 的优势是什么？" \
  --data-urlencode "top_k=5" \
  "http://127.0.0.1:8000/queries/rag" \
| jq '{answer, retrieved:(.top_chunks // [])}'
```

```
(base) [root@ecs-dbd1 ~]# curl -s -G -H "x-api-key: $KEY" \
>    --data-urlencode "workspace_id=$WS_ID" \
>    --data-urlencode "text=openGauss 的优势是什么? " \
>    --data-urlencode "top_k=5" \
>    "http://127.0.0.1:8000/queries/rag" \
> | jq '{answer, retrieved:(.top_chunks // [])}'
{
  "answer": "openGauss 的优势包括其企业级开源特性、与 PostgreSQL 的兼容性，以及具备高可用性和高性能。",
  "retrieved": [
    {
      "id": "d36d63d0-d55f-470c-b9b4-e208c3fd31e7",
      "score": 0.7228646348533775,
      "data_type": "string",
      "document_id": null
    },
    {
      "id": "4aedc4f1-b5ef-4424-bf4f-230701882acf",
      "score": 0.31803927749919614,
      "data_type": "string",
      "document_id": null
    },
    {
      "id": "9b245d7d-7420-4fba-9ac0-c3b1a1a2c58f",
      "score": 0.1085617521145224,
      "data_type": "string",
      "document_id": null
    },
    {
      "id": "73ee3834-b1b2-4aa1-a4bb-725d03fbeedf",
      "score": 0.10846599588685951,
      "data_type": "string",
      "document_id": null
    },
    {
      "id": "d0d6262b-b176-4f61-813d-15416536c096",
      "score": 0.0832668149443006,
      "data_type": "object",
      "document_id": null
    }
  ]
}
```

2. POST

```
jq -n --arg ws "$WS_ID" --arg txt "openGauss 的优势是什么? " --argjson top 5 \
  '{workspace_id:$ws, text:$txt, top_k:$top}' > /tmp/rag_req.json

curl -s -X POST \
  -H "x-api-key: $KEY" -H "Content-Type: application/json" \
  --data-binary @/tmp/rag_req.json \
  "http://127.0.0.1:8000/queries/rag" \
| tee /tmp/rag_resp.json \
| jq '{answer, retrieved: (.retrieval // [] | map({id, score, preview:
((.content|tostring)[0:80])}))}'
```

```
(base) [root@ecs-dbd1 ~]# curl -s -X POST -H "x-api-key: $KEY" \
>   -H "Content-Type: application/json" \
>   --data-binary @/tmp/rag_req.json \
>   "http://127.0.0.1:8000/queries/rag" \
> | jq '{answer, retrieved: (.top_chunks // [])}'
{
  "answer": "openGauss 的优势包括其企业级开源特性、与 PostgreSQL 的兼容性，以及高可用性和高性能。这使得它适合用于各种企业
级应用。",
  "retrieved": [
    {
      "id": "d36d63d0-d55f-470c-b9b4-e208c3fd31e7",
      "score": 0.7228679182328978,
      "data_type": "string",
      "document_id": null
    },
    {
      "id": "4aedc4f1-b5ef-4424-bf4f-230701882acf",
      "score": 0.31804382137614085,
      "data_type": "string",
      "document_id": null
    },
    {
      "id": "9b245d7d-7420-4fba-9ac0-c3b1a1a2c58f",
      "score": 0.10860657112143503,
      "data_type": "string",
      "document_id": null
    },
    {
      "id": "73ee3834-b1b2-4aa1-a4bb-725d03fbeedf",
      "score": 0.10851080417425106,
      "data_type": "string",
      "document_id": null
    },
    {
      "id": "d0d6262b-b176-4f61-813d-15416536c096",
      "score": 0.08330251679555153,
      "data_type": "object",
      "document_id": null
    }
  ]
}
```

3. 查列表

```
curl -s -H "x-api-key: $KEY" \
  "http://127.0.0.1:8000/queries?limit=5" | jq
```

```
(base) [root@ecs-dbd1 ~]# curl -s -H "x-api-key: $KEY" \
>    "http://127.0.0.1:8000/queries?limit=5" | jq
{
  "message": "ok",
  "status": "success",
  "count": 2,
  "queries": [
    {
      "id": "8151989b-3969-4573-93c0-c4cc45e89b09",
      "user_id": "12345678-1234-1234-1234-1234567890ab",
      "graph_id": null,
      "status": "completed",
      "name": "openGauss 的优势是什么？",
      "payload": {
        "request": {
          "workspace_id": "83a48b27-7ed9-4111-ad19-ca910e477653",
          "text": "openGauss 的优势是什么？",
          "top_k": 5,
          "filters": {}
        },
        "hits": [
          {
            "id": "d36d63d0-d55f-470c-b9b4-e208c3fd31e7",
            "score": 0.7228646348533775,
            "data_type": "string",
            "document_id": null
          },
          {
            "id": "4aedc4f1-b5ef-4424-bf4f-230701882acf",
            "score": 0.31803927749919614,
            "data_type": "string",
            "document_id": null
          },
          {
            "id": "9b245d7d-7420-4fba-9ac0-c3b1a1a2c58f",
            "score": 0.1085617521145224,
            "data_type": "string",
            "document_id": null
          },
          {
            "id": "73ee3834-b1b2-4aa1-a4bb-725d03fbeedf",
            "score": 0.10846599588685951,
            "data_type": "string",
            "document_id": null
          },
          {
            "id": "d0d6262b-b176-4f61-813d-15416536c096",
            "score": 0.0832668149443006,
            "data_type": "object",
            "document_id": null
          }
        ],
```

```
{
    "id": "178bb67b-053f-4093-adb3-06cf337fc004",
    "user_id": "12345678-1234-1234-1234-1234567890ab",
    "graph_id": null,
    "status": "completed",
    "name": "openGauss 的优势是什么？",
    "payload": {
      "request": {
        "workspace_id": "83a48b27-7ed9-4111-ad19-ca910e477653",
        "text": "openGauss 的优势是什么？",
        "top_k": 5,
        "filters": {}
      },
      "hits": [
        {
          "id": "d36d63d0-d55f-470c-b9b4-e208c3fd31e7",
          "score": 0.7228679182328978,
          "data_type": "string",
          "document_id": null
        },
        {
          "id": "4aedc4f1-b5ef-4424-bf4f-230701882acf",
          "score": 0.31804382137614085,
          "data_type": "string",
          "document_id": null
        },
        {
          "id": "9b245d7d-7420-4fba-9ac0-c3b1a1a2c58f",
          "score": 0.10860657112143503,
          "data_type": "string",
          "document_id": null
        },
        {
          "id": "73ee3834-b1b2-4aa1-a4bb-725d03fbeedf",
          "score": 0.10851080417425106,
          "data_type": "string",
          "document_id": null
        },
        {
          "id": "d0d6262b-b176-4f61-813d-15416536c096",
          "score": 0.08330251679555153,
          "data_type": "object",
          "document_id": null
        }
      ],
      "answer": "openGauss 的优势包括其企业级开源特性、与 PostgreSQL 的兼容性，以及高可用性和高性能。这使得它适合用于各种企业级应用。"
    },
    "created_at": "2025-09-10T03:42:31.654090+00:00",
    "updated_at": "2025-09-10T03:42:31.654090+00:00"
  }
]
}
```