

结构化JSON

14:57:06 [http-nio-9080-exec-2] DEBUG
c.t.s.h.s.w.s.i.WorkflowServiceImpl 67 - RuleSqlParser
result:

```
{
  "queryText": "超音速对比所有人访问次数",
  "chatId": 1,
  "dataSetIds": [],
  "modelIdToDataSetIds": { "1": [1], "2": [1], "3": [1], "4":
[2] },
  "user": {
    "id": 6,
    "name": "admin",
    "displayName": "admin",
    "email": "null",
    "isAdmin": 1,
    "superAdmin": true
  },
  "saveAnswer": true,
  "text2SQLType": "RULE_AND_LLM",
  "candidateQueries": [
    {
      "parseInfo": {
        "queryMode": "METRIC_GROUPBY",
        "dataSet": {
          "dataSet": 1,
          "dataSetName": "超音数数据集",
          "id": 1,
          "name": "超音数数据集",
          "bizName": "s2",
          "type": "DATASET",
          "order": 0.0,
          "isTag": 0
        },
      },
      "metrics": [
        {
```

```

    "dataSet": 1,
    "dataSetName": "超音数数据集",
    "model": 2,
    "id": 1,
    "name": "访问次数",
    "bizName": "pv",
    "useCnt": 0,
    "type": "METRIC",
    "alias": [],
    "relatedSchemaElements": [
      { "dimensionId": 1, "necessary": false },
      { "dimensionId": 2, "necessary": false }
    ],
    "defaultAgg": "SUM",
    "order": 0.0,
    "isTag": 0,
    "description": "一段时间内用户的访问次数"
  },
  {
    "dataSet": 1,
    "dataSetName": "超音数数据集",
    "model": 2,
    "id": 4,
    "name": "人均访问次数",
    "bizName": "pv_avg",
    "useCnt": 0,
    "type": "METRIC",
    "alias": ["平均访问次数"],
    "relatedSchemaElements": [{ "dimensionId": 1, "necessary": false }],
    "defaultAgg": "",
    "order": 0.002841870251016343,
    "isTag": 0,
    "description": "每个用户平均访问的次数"
  },
  {
    "dataSet": 1,
    "dataSetName": "超音数数据集",
    "model": 2,
    "id": 3,

```

```

        "name": "访问用户数",
        "bizName": "uv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["UV", "访问人数"],
        "relatedSchemaElements": [{ "dimensionId": 1, "necessary": false }],
        "defaultAgg": "count",
        "order": 0.004658766300494777,
        "isTag": 0,
        "description": "访问的用户个数"
    }
],
"dimensions": [
    {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 1,
        "id": 1,
        "name": "部门",
        "bizName": "department",
        "useCnt": 0,
        "type": "DIMENSION",
        "alias": [],
        "schemaValueMaps": [],
        "order": 0.0,
        "isTag": 1,
        "description": ""
    }
],
"aggType": "NONE",
"filterType": "AND",
"dimensionFilters": [],
"metricFilters": [],
"orders": [],
"dateInfo": {
    "dateMode": "RECENT",
    "startDate": "2024-06-08",
    "endDate": "2024-07-07",
    "dateList": [

```

```
        "2024-06-08",
        "2024-06-09",
        "2024-06-10",
        "2024-06-11",
        "2024-06-12",
        "2024-06-13",
        "2024-06-14",
        "2024-06-15",
        "2024-06-16",
        "2024-06-17",
        "2024-06-18",
        "2024-06-19",
        "2024-06-20",
        "2024-06-21",
        "2024-06-22",
        "2024-06-23",
        "2024-06-24",
        "2024-06-25",
        "2024-06-26",
        "2024-06-27",
        "2024-06-28",
        "2024-06-29",
        "2024-06-30",
        "2024-07-01",
        "2024-07-02",
        "2024-07-03",
        "2024-07-04",
        "2024-07-05",
        "2024-07-06",
        "2024-07-07"
    ],
    "unit": 30,
    "period": "DAY",
    "detectWord": "近30天",
    "groupByDate": false,
    "inherited": true,
    "groupByTimeDimension": "sys_imp_date"
},
"limit": 365,
"score": 6.0,
```

```

"elementMatches": [
  {
    "element": {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "model": 2,
      "id": 1,
      "name": "访问次数",
      "bizName": "pv",
      "useCnt": 0,
      "type": "METRIC",
      "alias": [],
      "relatedSchemaElements": [
        { "dimensionId": 1, "necessary": false },
        { "dimensionId": 2, "necessary": false }
      ],
      "defaultAgg": "SUM",
      "order": 0.0,
      "isTag": 0,
      "description": "一段时间内用户的访问次数"
    },
    "similarity": 1.0,
    "detectWord": "访问次数",
    "word": "访问次数",
    "frequency": 100000,
    "inherited": false
  },
  {
    "element": {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "model": 2,
      "id": 4,
      "name": "人均访问次数",
      "bizName": "pv_avg",
      "useCnt": 0,
      "type": "METRIC",
      "alias": ["平均访问次数"],
      "relatedSchemaElements": [
        { "dimensionId": 1, "necessary": false }
      ]
    }
  }
]

```

```

    ],
    "defaultAgg": "",
    "order": 0.002841870251016343,
    "isTag": 0,
    "description": "每个用户平均访问的次数"
  },
  "similarity": 0.9971581297489837,
  "detectWord": "人访问次数",
  "word": "人均访问次数",
  "frequency": 100000,
  "inherited": false
},
{
  "element": {
    "dataSet": 1,
    "dataSetName": "超音数数据集",
    "model": 2,
    "id": 3,
    "name": "访问用户数",
    "bizName": "uv",
    "useCnt": 0,
    "type": "METRIC",
    "alias": ["UV", "访问人数"],
    "relatedSchemaElements": [
      { "dimensionId": 1, "necessary": false }
    ],
    "defaultAgg": "count",
    "order": 0.004658766300494777,
    "isTag": 0,
    "description": "访问的用户个数"
  },
  "similarity": 0.9953412336995052,
  "detectWord": "访问次数",
  "word": "访问用户数",
  "frequency": 100000,
  "inherited": false
},
{
  "element": {
    "dataSet": 1,

```

```

        "dataSetName": "超音数数据集",
        "model": 1,
        "id": 1,
        "name": "部门",
        "bizName": "department",
        "useCnt": 0,
        "type": "DIMENSION",
        "alias": [],
        "schemaValueMaps": [],
        "order": 0.0,
        "isTag": 1,
        "description": ""
    },
    "similarity": 1.0,
    "detectWord": "部门",
    "word": "部门",
    "frequency": 100000,
    "inherited": true
}
],
"properties": {},
"sqlInfo": {},
"sqlEvaluation": {},
"queryType": "ID",
"dataSetId": 1
},
"queryMode": "METRIC_GROUPBY"
},
{
    "parseInfo": {
        "queryMode": "METRIC_MODEL",
        "dataSet": {
            "dataSet": 1,
            "dataSetName": "超音数数据集",
            "id": 1,
            "name": "超音数数据集",
            "bizName": "s2",
            "type": "DATASET",
            "order": 0.0,
            "isTag": 0
        }
    }
}

```

```

    },
    "metrics": [
      {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 1,
        "name": "访问次数",
        "bizName": "pv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": [],
        "relatedSchemaElements": [
          { "dimensionId": 1, "necessary": false },
          { "dimensionId": 2, "necessary": false }
        ],
        "defaultAgg": "SUM",
        "order": 0.0,
        "isTag": 0,
        "description": "一段时间内用户的访问次数"
      },
      {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 4,
        "name": "人均访问次数",
        "bizName": "pv_avg",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["平均访问次数"],
        "relatedSchemaElements": [{ "dimensionId": 1, "necessary": false }],
        "defaultAgg": "",
        "order": 0.002841870251016343,
        "isTag": 0,
        "description": "每个用户平均访问的次数"
      },
      {
        "dataSet": 1,

```



```

        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 3,
        "name": "访问用户数",
        "bizName": "uv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["UV", "访问人数"],
        "relatedSchemaElements": [{ "dimensionId": 1, "necessary": false }],
        "defaultAgg": "count",
        "order": 0.004658766300494777,
        "isTag": 0,
        "description": "访问的用户个数"
    }
],
"dimensions": [],
"aggType": "NONE",
"filterType": "AND",
"dimensionFilters": [],
"metricFilters": [],
"orders": [],
"dateInfo": {
    "dateMode": "RECENT",
    "startDate": "2024-06-08",
    "endDate": "2024-07-07",
    "dateList": [
        "2024-06-08",
        "2024-06-09",
        "2024-06-10",
        "2024-06-11",
        "2024-06-12",
        "2024-06-13",
        "2024-06-14",
        "2024-06-15",
        "2024-06-16",
        "2024-06-17",
        "2024-06-18",
        "2024-06-19",
        "2024-06-20",
    ]
}

```

```
        "2024-06-21",
        "2024-06-22",
        "2024-06-23",
        "2024-06-24",
        "2024-06-25",
        "2024-06-26",
        "2024-06-27",
        "2024-06-28",
        "2024-06-29",
        "2024-06-30",
        "2024-07-01",
        "2024-07-02",
        "2024-07-03",
        "2024-07-04",
        "2024-07-05",
        "2024-07-06",
        "2024-07-07"
    ],
    "unit": 30,
    "period": "DAY",
    "detectWord": "近30天",
    "groupByDate": false,
    "inherited": true,
    "groupByTimeDimension": "sys_imp_date"
},
"limit": 365,
"score": 4.0,
"elementMatches": [
    {
        "element": {
            "dataSet": 1,
            "dataSetName": "超音数数据集",
            "model": 2,
            "id": 1,
            "name": "访问次数",
            "bizName": "pv",
            "useCnt": 0,
            "type": "METRIC",
            "alias": [],
            "relatedSchemaElements": [
```

```

        { "dimensionId": 1, "necessary": false },
        { "dimensionId": 2, "necessary": false }
    ],
    "defaultAgg": "SUM",
    "order": 0.0,
    "isTag": 0,
    "description": "一段时间内用户的访问次数"
},
"similarity": 1.0,
"detectWord": "访问次数",
"word": "访问次数",
"frequency": 100000,
"inherited": false
},
{
    "element": {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 4,
        "name": "人均访问次数",
        "bizName": "pv_avg",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["平均访问次数"],
        "relatedSchemaElements": [
            { "dimensionId": 1, "necessary": false }
        ],
        "defaultAgg": "",
        "order": 0.002841870251016343,
        "isTag": 0,
        "description": "每个用户平均访问的次数"
    },
    "similarity": 0.9971581297489837,
    "detectWord": "人访问次数",
    "word": "人均访问次数",
    "frequency": 100000,
    "inherited": false
},
{

```

```

        "element": {
            "dataSet": 1,
            "dataSetName": "超音数数据集",
            "model": 2,
            "id": 3,
            "name": "访问用户数",
            "bizName": "uv",
            "useCnt": 0,
            "type": "METRIC",
            "alias": ["UV", "访问人数"],
            "relatedSchemaElements": [
                { "dimensionId": 1, "necessary": false }
            ],
            "defaultAgg": "count",
            "order": 0.004658766300494777,
            "isTag": 0,
            "description": "访问的用户个数"
        },
        "similarity": 0.9953412336995052,
        "detectWord": "访问次数",
        "word": "访问用户数",
        "frequency": 100000,
        "inherited": false
    }
],
    "properties": {},
    "sqlInfo": {},
    "sqlEvaluation": {},
    "queryType": "ID",
    "dataSetId": 1
},
    "queryMode": "METRIC_MODEL"
}
],
    "mapInfo": {
        "dataSetElementMatches": {
            "1": [
                {
                    "element": {
                        "dataSet": 1,

```

```

    "dataSetName": "超音数数据集",
    "model": 2,
    "id": 1,
    "name": "访问次数",
    "bizName": "pv",
    "useCnt": 0,
    "type": "METRIC",
    "alias": [],
    "relatedSchemaElements": [
      { "dimensionId": 1, "necessary": false },
      { "dimensionId": 2, "necessary": false }
    ],
    "defaultAgg": "SUM",
    "order": 0.0,
    "isTag": 0,
    "description": "一段时间内用户的访问次数"
  },
  "similarity": 1.0,
  "detectWord": "访问次数",
  "word": "访问次数",
  "frequency": 100000,
  "inherited": false
},
{
  "element": {
    "dataSet": 1,
    "dataSetName": "超音数数据集",
    "model": 2,
    "id": 4,
    "name": "人均访问次数",
    "bizName": "pv_avg",
    "useCnt": 0,
    "type": "METRIC",
    "alias": ["平均访问次数"],
    "relatedSchemaElements": [{ "dimensionId": 1, "necessary": false }],
    "defaultAgg": "",
    "order": 0.002841870251016343,
    "isTag": 0,
    "description": "每个用户平均访问的次数"
  }
}

```

```

    },
    "similarity": 0.9971581297489837,
    "detectWord": "人访问次数",
    "word": "人均访问次数",
    "frequency": 100000,
    "inherited": false
  },
  {
    "element": {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "model": 2,
      "id": 3,
      "name": "访问用户数",
      "bizName": "uv",
      "useCnt": 0,
      "type": "METRIC",
      "alias": ["UV", "访问人数"],
      "relatedSchemaElements": [{ "dimensionId": 1, "necessary": false }],
      "defaultAgg": "count",
      "order": 0.004658766300494777,
      "isTag": 0,
      "description": "访问的用户个数"
    },
    "similarity": 0.9953412336995052,
    "detectWord": "访问次数",
    "word": "访问用户数",
    "frequency": 100000,
    "inherited": false
  },
  {
    "element": {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "model": 1,
      "id": 1,
      "name": "部门",
      "bizName": "department",
      "useCnt": 0,

```

```

        "type": "DIMENSION",
        "alias": [],
        "schemaValueMaps": [],
        "order": 0.0,
        "isTag": 1,
        "description": ""
    },
    "similarity": 1.0,
    "detectWord": "部门",
    "word": "部门",
    "frequency": 100000,
    "inherited": true
}
]
},
"matchedDataSetInfos": [1]
},
"mapModeEnum": "STRICT",
"queryDataType": "ALL",
"llmConfig": {
    "provider": "OPEN_AI",
    "baseUrl": "https://api.openai.com/v1",
    "apiKey": "kGA4i6jY7QqG5xH+S0N++SzVWVKyAHes0q7NpoFVI7pgU1
4eb1AM2960FiLJTntDVfpWDpfu7o4vDG2C2yWiSA==",
    "modelName": "gpt-3.5-turbo",
    "temperature": 0.0,
    "timeOut": 60
},
"exemplars": []
}

```

解释：

这个JSON数据包含了一个查询请求的详细信息，其中包括用户信息、查询文本、相关的数据集和模型、LLM（大语言模型）配置等。以下是对这个JSON的逐项详细解释：

顶层结构

```
{
  "queryText": "超音速对比所有人访问次数",
  "chatId": 1,
  "dataSetIds": [],
  "modelIdToDataSetIds": { "1": [1], "2": [1], "3": [1], "4":
[2] },
  "user": {...},
  "saveAnswer": true,
  "text2SQLType": "RULE_AND_LLM",
  "candidateQueries": [...],
  "mapInfo": {...},
  "mapModeEnum": "STRICT",
  "queryDataType": "ALL",
  "llmConfig": {...},
  "exemplars": []
}
```

1. **queryText** : "超音速对比所有人访问次数"
 - 这是用户输入的查询文本。用户希望查询“超音速对比所有人访问次数”的数据。
2. **chatId** : 1
 - 这是对话的唯一标识符，表示这个查询所属的对话ID。
3. **dataSetIds** : []
 - 这是一个空数组，通常用于指定查询涉及的数据集ID。这里为空，可能意味着查询时还没有指定具体的数据集。
4. **modelIdToDataSetIds** : { "1": [1], "2": [1], "3": [1], "4": [2] }
 - 这是一个映射关系，表示模型ID到数据集ID的映射关系。例如，模型ID为1、2、3的都对应数据集ID为1，而模型ID为4的对应数据集ID为2。
5. **user** :

```
{
  "id": 6,
  "name": "admin",
  "displayName": "admin",
  "email": "null",
```



```
"isAdmin": 1,  
"superAdmin": true}
```

- 这个对象包含了执行查询的用户信息：
 - `id: 6` 表示用户ID。
 - `name: "admin"` 是用户名。
 - `displayName: "admin"` 是显示名称。
 - `email: "null"` 表示用户邮箱地址。
 - `isAdmin: 1` 表示该用户是管理员。
 - `superAdmin: true` 表示该用户是超级管理员。

6. `saveAnswer: true`

- 这表示查询的结果是否需要保存。`true` 表示结果将被保存。

7. `text2SQLType: "RULE_AND_LLM"`

- 这是用于将自然语言查询转换为SQL的策略类型，`RULE_AND_LLM` 表示该过程结合了规则和大语言模型（LLM）。

8. `candidateQueries:`

- 这是一个数组，包含多个候选查询，每个候选查询可能代表解析查询文本后的不同SQL查询。

9. `mapInfo:`

- 这是一个对象，包含数据集元素匹配的信息和匹配的数据信息，详细解释见下文。

10. `mapModeEnum: "STRICT"`

- 这是映射模式，`STRICT` 表示严格模式，可能意味着对查询解析的精确性要求较高。

11. `queryDataType: "ALL"`

- 这表示查询的数据类型，`ALL` 表示查询包括所有类型的数据。

12. `llmConfig:`

- 这是一个对象，包含与大语言模型（LLM）相关的配置信息，详细解释见下文。

13. `exemplars: []`

- 这是一个空数组，通常用于存储查询示例或训练数据。在此例中，没有提供示例。

candidateQueries 数组中的项

这是系统生成的多个候选查询，系统会选择其中一个或多个进行执行。

第一个候选查询

```
{
  "parseInfo": {
    "queryMode": "METRIC_GROUPBY",
    "dataSet": {...},
    "metrics": [...],
    "dimensions": [...],
    "aggType": "NONE",
    "filterType": "AND",
    "dimensionFilters": [],
    "metricFilters": [],
    "orders": [],
    "dateInfo": {...},
    "limit": 365,
    "score": 6.0,
    "elementMatches": [...],
    "properties": {},
    "sqlInfo": {},
    "sqlEvaluation": {},
    "queryType": "ID",
    "dataSetId": 1
  },
  "queryMode": "METRIC_GROUPBY"
}
```

1. `parseInfo.queryMode` : "METRIC_GROUPBY"
 - 表示查询模式为按维度分组进行度量计算。
2. `parseInfo.dataSet` :
 - 包含与查询相关的数据集信息，详细解释见下文。
3. `parseInfo.metrics` :
 - 包含查询涉及的指标（metrics）信息，详细解释见下文。
4. `parseInfo.dimensions` :

- 包含查询涉及的维度（dimensions）信息，详细解释见下文。
5. `parseInfo.aggType` : "NONE"
 - 表示查询没有进行额外的聚合操作。
 6. `parseInfo.filterType` : "AND"
 - 表示查询过滤条件使用的是 AND 逻辑。
 7. `parseInfo.dimensionFilters` : []
 - 这是一个空数组，表示没有指定维度过滤条件。
 8. `parseInfo.metricFilters` : []
 - 这是一个空数组，表示没有指定指标过滤条件。
 9. `parseInfo.orders` : []
 - 这是一个空数组，表示没有指定查询结果的排序顺序。
 10. `parseInfo.dateInfo` :
 - 包含查询涉及的时间范围信息，详细解释见下文。
 11. `parseInfo.limit` : 365
 - 表示查询结果的数量限制为365条记录。
 12. `parseInfo.score` : 6.0
 - 这是系统对该解析结果的评分，分数越高表示该解析结果与原始查询的匹配度越高。
 13. `parseInfo.elementMatches` :
 - 包含查询中识别出的元素匹配信息，详细解释见下文。
 14. `parseInfo.properties` : {}
 - 这是一个空对象，表示没有额外的属性信息。
 15. `parseInfo.sqlInfo` : {}
 - 这是一个空对象，表示还没有生成SQL查询语句。
 16. `parseInfo.sqlEvaluation` : {}
 - 这是一个空对象，表示没有进行SQL评估。
 17. `parseInfo.queryType` : "ID"
 - 这表示查询的类型，这里是ID类型查询，可能与系统内部逻辑相关。
 18. `parseInfo.dataSetId` : 1

- 表示使用的数据集ID为1。

19. `queryMode`: "METRIC_GROUPBY"

- 这与 `parseInfo.queryMode` 相同，表示查询模式为按维度分组进行度量计算。

第二个候选查询

```
{
  "parseInfo": {
    "queryMode": "METRIC_MODEL",
    "dataSet": {...},
    "metrics": [...],
    "dimensions": [],
    "aggType": "NONE",
    "filterType": "AND",
    "dimensionFilters": [],
    "metricFilters": [],
    "orders": [],
    "dateInfo": {...},
    "limit": 365,
    "score": 4.0,
    "elementMatches": [...],
    "properties": {},
    "sqlInfo": {},
    "sqlEvaluation": {},
    "queryType": "ID",
    "dataSetId": 1
  },
  "queryMode": "METRIC_MODEL"
}
```

与第一个候选查询类似，但 `queryMode` 为 "METRIC_MODEL"，可能表示一种不同的查询模式，聚焦于模型相关的度量计算。评分为 4.0，比第一个候选查询略低。

`mapInfo` 对象

```
{
  "dataSetElementMatches": {
    "1": [
      {...},
    ]
  }
}
```

```

        {...},
        {...},
        {...}
    ]
},
"matchedDataSetInfos": [1]
}

```

1. **dataSetElementMatches** :

- 这是一个映射，键是数据集ID，值是该数据集中匹配的元素列表。例如，数据集ID为 **1** 的数据集中，有多个匹配的元素（如指标和维度）。

2. **matchedDataSetInfos** : **[1]**

- 这是一个数组，列出了匹配的数据集ID。在这个例子中，只有数据集ID为1的被匹配。

llmConfig 对象

```

{
  "provider": "OPEN_AI",
  "baseUrl": "https://api.openai.com/v1",
  "apiKey": "kGA4i6jY7QqG5xH+S0N++SzVWVKyAHes0q7NpoFVI7pgU14eb1AM2960FiLJTntDVfpWDpfu7o4vDG2C2yWiSA==",
  "modelName": "gpt-3.5-turbo",
  "temperature": 0.0,
  "timeOut": 60
}

```

1. **provider** : **"OPEN_AI"**

- 表示使用的LLM提供商是OpenAI。

2. **baseUrl** : **"https://api.openai.com/v1"**

- 这是LLM API的基础URL。

3. **apiKey** :

"kGA4i6jY7QqG5xH+S0N++SzVWVKyAHes0q7NpoFVI7pgU14eb1AM2960FiLJTntDVfpWDpfu7o4vDG2yWiSA=="

- 这是调用API所需的密钥。

4. **modelName** : **"gpt-3.5-turbo"**

- 这是使用的模型名称，表示使用了OpenAI的GPT-3.5 Turbo模型。

5. `temperature: 0.0`
 - 这是生成文本的“温度”参数，控制输出的随机性。`0.0` 表示输出结果更加确定性，较少随机性。
6. `timeOut: 60`
 - 这是API请求的超时时间，单位为秒。

关键点总结

1. **查询模式**：该JSON定义了两个候选查询模式——`METRIC_GROUPBY` 和 `METRIC_MODEL`，分别处理不同的查询逻辑。
2. **用户信息**：请求由用户 `admin` 发起，且为管理员身份。
3. **数据集匹配**：查询的文本与数据集中的指标和维度进行了匹配，并选择了合适的模式生成候选SQL查询。
4. **LLM配置**：使用了OpenAI的GPT-3.5 Turbo模型，结合规则和LLM来处理自然语言查询并生成SQL。

14:57:19 [http-nio-9080-exec-2] DEBUG
c.t.s.h.s.w.s.i.WorkflowServiceImpl 67 - LLMSqlParser
result:

```
{
  "queryText": "超音速对比所有人访问次数",
  "chatId": 1,
  "dataSetIds": [],
  "modelIdToDataSetIds": { "1": [1], "2": [1], "3": [1], "4": [
  "user": {
    "id": 6,
    "name": "admin",
    "displayName": "admin",
    "email": "null",
    "isAdmin": 1,
    "superAdmin": true
  },
  "saveAnswer": true,
  "text2SQLType": "RULE_AND_LLM",
  "candidateQueries": [
```

```

{
  "parseInfo": {
    "queryMode": "LLM_S2SQL",
    "dataSet": {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "id": 1,
      "name": "超音数数据集",
      "bizName": "s2",
      "type": "DATASET",
      "order": 0.0,
      "isTag": 0
    },
    "metrics": [],
    "dimensions": [],
    "aggType": "NONE",
    "filterType": "AND",
    "dimensionFilters": [],
    "metricFilters": [],
    "orders": [],
    "score": 24.0,
    "elementMatches": [
      {
        "element": {
          "dataSet": 1,
          "dataSetName": "超音数数据集",
          "model": 2,
          "id": 1,
          "name": "访问次数",
          "bizName": "pv",
          "useCnt": 0,
          "type": "METRIC",
          "alias": [],
          "relatedSchemaElements": [
            { "dimensionId": 1, "necessary": false },
            { "dimensionId": 2, "necessary": false }
          ],
          "defaultAgg": "SUM",
          "order": 0.0,
          "isTag": 0,

```

```

        "description": "一段时间内用户的访问次数"
    },
    "similarity": 1.0,
    "detectWord": "访问次数",
    "word": "访问次数",
    "frequency": 100000,
    "inherited": false
},
{
    "element": {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 4,
        "name": "人均访问次数",
        "bizName": "pv_avg",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["平均访问次数"],
        "relatedSchemaElements": [
            { "dimensionId": 1, "necessary": false }
        ],
        "defaultAgg": "",
        "order": 0.002841870251016343,
        "isTag": 0,
        "description": "每个用户平均访问的次数"
    },
    "similarity": 0.9971581297489837,
    "detectWord": "人访问次数",
    "word": "人均访问次数",
    "frequency": 100000,
    "inherited": false
},
{
    "element": {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 3,
        "name": "访问用户数",

```



```

        "bizName": "uv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["UV", "访问人数"],
        "relatedSchemaElements": [
            { "dimensionId": 1, "necessary": false }
        ],
        "defaultAgg": "count",
        "order": 0.004658766300494777,
        "isTag": 0,
        "description": "访问的用户个数"
    },
    "similarity": 0.9953412336995052,
    "detectWord": "访问次数",
    "word": "访问用户数",
    "frequency": 100000,
    "inherited": false
},
{
    "element": {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 1,
        "id": 1,
        "name": "部门",
        "bizName": "department",
        "useCnt": 0,
        "type": "DIMENSION",
        "alias": [],
        "schemaValueMaps": [],
        "order": 0.0,
        "isTag": 1,
        "description": ""
    },
    "similarity": 1.0,
    "detectWord": "部门",
    "word": "部门",
    "frequency": 100000,
    "inherited": true
}

```

```

],
"properties": {
  "type": "internal",
  "CONTEXT": {
    "dataSetId": 1,
    "llmReq": {
      "queryText": "超音速对比所有人访问次数",
      "filterCondition": {},
      "schema": {
        "domainName": "超音数数据集",
        "dataSetName": "超音数数据集",
        "dataSetId": 1,
        "fieldNameList": [
          "访问用户数",
          "用户",
          "部门",
          "访问次数",
          "人均访问次数",
          "停留时长",
          "页面",
          "数据日期"
        ],
      },
    },
  },
  "metrics": [
    {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "model": 2,
      "id": 1,
      "name": "访问次数",
      "bizName": "pv",
      "useCnt": 0,
      "type": "METRIC",
      "alias": [],
      "relatedSchemaElements": [
        { "dimensionId": 1, "necessary": false },
        { "dimensionId": 2, "necessary": false }
      ],
      "defaultAgg": "SUM",
      "order": 0.0,
      "isTag": 0,
    }
  ]
}

```

```

        "description": "一段时间内用户的访问次数"
    },
    {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 4,
        "name": "人均访问次数",
        "bizName": "pv_avg",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["平均访问次数"],
        "relatedSchemaElements": [
            { "dimensionId": 1, "necessary": false }
        ],
        "defaultAgg": "",
        "order": 0.002841870251016343,
        "isTag": 0,
        "description": "每个用户平均访问的次数"
    },
    {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 3,
        "name": "访问用户数",
        "bizName": "uv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["UV", "访问人数"],
        "relatedSchemaElements": [
            { "dimensionId": 1, "necessary": false }
        ],
        "defaultAgg": "count",
        "order": 0.004658766300494777,
        "isTag": 0,
        "description": "访问的用户个数"
    }
],
"dimensions": [

```

```

        {
            "dataSet": 1,
            "dataSetName": "超音数数据集",
            "model": 1,
            "id": 1,
            "name": "部门",
            "bizName": "department",
            "useCnt": 0,
            "type": "DIMENSION",
            "alias": [],
            "schemaValueMaps": [],
            "order": 0.0,
            "isTag": 1,
            "description": ""
        }
    ],
    "terms": []
},
"linking": [],
"currentDate": "2024-08-07",
"priorExts": "",
"sqlGenType": "1_pass_self_consistency",
"llmConfig": {
    "provider": "OPEN_AI",
    "baseUrl": "https://api.openai.com/v1",
    "apiKey": "kGA4i6jY7QqG5xH+S0N++SzVWVKyAHes0q7N",
    "modelName": "gpt-3.5-turbo",
    "temperature": 0.0,
    "timeOut": 60
},
"exemplars": []
},
"llmResp": {
    "query": "超音速对比所有人访问次数",
    "modelName": "超音数数据集",
    "dbSchema": "Table: 超音数数据集, Metrics: [访问次数",
    "sqlOutput": "SELECT SUM(访问次数) FROM 超音数数据集",
    "sqlRespMap": {
        "SELECT SUM(访问次数) FROM 超音数数据集": {
            "sqlWeight": 1.0,

```

```

        "fewShots": [
            {
                "question": "今年以来发行的歌曲中，有哪些在近7天",
                "dbSchema": "Table: 歌曲库, Columns = [\\"发",
                "sql": "SELECT 歌曲名 FROM 歌曲库 WHERE dat",
            },
            {
                "question": "超音数近半年每个月的平均访问次数",
                "dbSchema": "Table: 超音数产品, Columns = [",
                "sql": "SELECT MONTH(数据日期), AVG(访问次数",
            },
            {
                "question": "我想要近半年签约的播放量前十的歌手",
                "dbSchema": "Table: 艺人库, Columns = [\\"播",
                "sql": "SELECT 歌手名 FROM 艺人库 WHERE dat",
            },
            {
                "question": "播放量大于1万的歌曲有多少（补充信",
                "dbSchema": "Table: 歌曲库, Columns = [\\"歌",
                "sql": "SELECT 歌曲名 FROM 歌曲库 WHERE 结算",
            },
            {
                "question": "超音数访问次数大于1k的部门是哪些",
                "dbSchema": "Table: 超音数产品, Columns = [",
                "sql": "SELECT 部门 FROM 超音数产品 WHERE 访",
            }
        ]
    },
    "linkingValues": []
}

},
"sqlInfo": { "s2SQL": "SELECT SUM(访问次数) FROM 超音数数",
"sqlEvaluation": {},
"queryType": "ID",
"dataSetId": 1
},
"queryMode": "LLM_S2SQL"
},

```

```

{
  "parseInfo": {
    "queryMode": "METRIC_GROUPBY",
    "dataSet": {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "id": 1,
      "name": "超音数数据集",
      "bizName": "s2",
      "type": "DATASET",
      "order": 0.0,
      "isTag": 0
    },
    "metrics": [
      {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 1,
        "name": "访问次数",
        "bizName": "pv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": [],
        "relatedSchemaElements": [
          { "dimensionId": 1, "necessary": false },
          { "dimensionId": 2, "necessary": false }
        ],
        "defaultAgg": "SUM",
        "order": 0.0,
        "isTag": 0,
        "description": "一段时间内用户的访问次数"
      },
      {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 4,
        "name": "人均访问次数",
        "bizName": "pv_avg",

```

```

        "useCnt": 0,
        "type": "METRIC",
        "alias": ["平均访问次数"],
        "relatedSchemaElements": [{ "dimensionId": 1, "nece
        "defaultAgg": "",
        "order": 0.002841870251016343,
        "isTag": 0,
        "description": "每个用户平均访问的次数"
    },
    {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 3,
        "name": "访问用户数",
        "bizName": "uv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["UV", "访问人数"],
        "relatedSchemaElements": [{ "dimensionId": 1, "nece
        "defaultAgg": "count",
        "order": 0.004658766300494777,
        "isTag": 0,
        "description": "访问的用户个数"
    }
],
"dimensions": [
    {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 1,
        "id": 1,
        "name": "部门",
        "bizName": "department",
        "useCnt": 0,
        "type": "DIMENSION",
        "alias": [],
        "schemaValueMaps": [],
        "order": 0.0,
        "isTag": 1,

```

```
        "description": ""
    }
],
"aggType": "NONE",
"filterType": "AND",
"dimensionFilters": [],
"metricFilters": [],
"orders": [],
"dateInfo": {
    "dateMode": "RECENT",
    "startDate": "2024-06-08",
    "endDate": "2024-07-07",
    "dateList": [
        "2024-06-08",
        "2024-06-09",
        "2024-06-10",
        "2024-06-11",
        "2024-06-12",
        "2024-06-13",
        "2024-06-14",
        "2024-06-15",
        "2024-06-16",
        "2024-06-17",
        "2024-06-18",
        "2024-06-19",
        "2024-06-20",
        "2024-06-21",
        "2024-06-22",
        "2024-06-23",
        "2024-06-24",
        "2024-06-25",
        "2024-06-26",
        "2024-06-27",
        "2024-06-28",
        "2024-06-29",
        "2024-06-30",
        "2024-07-01",
        "2024-07-02",
        "2024-07-03",
        "2024-07-04",
```



```

        "2024-07-05",
        "2024-07-06",
        "2024-07-07"
    ],
    "unit": 30,
    "period": "DAY",
    "detectWord": "近30天",
    "groupByDate": false,
    "inherited": true,
    "groupByTimeDimension": "sys_imp_date"
},
"limit": 365,
"score": 6.0,
"elementMatches": [
    {
        "element": {
            "dataSet": 1,
            "dataSetName": "超音数数据集",
            "model": 2,
            "id": 1,
            "name": "访问次数",
            "bizName": "pv",
            "useCnt": 0,
            "type": "METRIC",
            "alias": [],
            "relatedSchemaElements": [
                { "dimensionId": 1, "necessary": false },
                { "dimensionId": 2, "necessary": false }
            ],
            "defaultAgg": "SUM",
            "order": 0.0,
            "isTag": 0,
            "description": "一段时间内用户的访问次数"
        },
        "similarity": 1.0,
        "detectWord": "访问次数",
        "word": "访问次数",
        "frequency": 100000,
        "inherited": false
    },

```

```

{
  "element": {
    "dataSet": 1,
    "dataSetName": "超音数数据集",
    "model": 2,
    "id": 4,
    "name": "人均访问次数",
    "bizName": "pv_avg",
    "useCnt": 0,
    "type": "METRIC",
    "alias": ["平均访问次数"],
    "relatedSchemaElements": [
      { "dimensionId": 1, "necessary": false }
    ],
    "defaultAgg": "",
    "order": 0.002841870251016343,
    "isTag": 0,
    "description": "每个用户平均访问的次数"
  },
  "similarity": 0.9971581297489837,
  "detectWord": "人访问次数",
  "word": "人均访问次数",
  "frequency": 100000,
  "inherited": false
},
{
  "element": {
    "dataSet": 1,
    "dataSetName": "超音数数据集",
    "model": 2,
    "id": 3,
    "name": "访问用户数",
    "bizName": "uv",
    "useCnt": 0,
    "type": "METRIC",
    "alias": ["UV", "访问人数"],
    "relatedSchemaElements": [
      { "dimensionId": 1, "necessary": false }
    ],
    "defaultAgg": "count",
  }
}

```

```

        "order": 0.004658766300494777,
        "isTag": 0,
        "description": "访问的用户个数"
    },
    "similarity": 0.9953412336995052,
    "detectWord": "访问次数",
    "word": "访问用户数",
    "frequency": 100000,
    "inherited": false
},
{
    "element": {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 1,
        "id": 1,
        "name": "部门",
        "bizName": "department",
        "useCnt": 0,
        "type": "DIMENSION",
        "alias": [],
        "schemaValueMaps": [],
        "order": 0.0,
        "isTag": 1,
        "description": ""
    },
    "similarity": 1.0,
    "detectWord": "部门",
    "word": "部门",
    "frequency": 100000,
    "inherited": true
}
],
"properties": {},
"sqlInfo": {},
"sqlEvaluation": {},
"queryType": "ID",
"dataSetId": 1
},
"queryMode": "METRIC_GROUPBY"

```

```

},
{
  "parseInfo": {
    "queryMode": "METRIC_MODEL",
    "dataSet": {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "id": 1,
      "name": "超音数数据集",
      "bizName": "s2",
      "type": "DATASET",
      "order": 0.0,
      "isTag": 0
    },
    "metrics": [
      {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 1,
        "name": "访问次数",
        "bizName": "pv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": [],
        "relatedSchemaElements": [
          { "dimensionId": 1, "necessary": false },
          { "dimensionId": 2, "necessary": false }
        ],
        "defaultAgg": "SUM",
        "order": 0.0,
        "isTag": 0,
        "description": "一段时间内用户的访问次数"
      },
      {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 4,
        "name": "人均访问次数",

```

```

        "bizName": "pv_avg",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["平均访问次数"],
        "relatedSchemaElements": [{ "dimensionId": 1, "nece
        "defaultAgg": "",
        "order": 0.002841870251016343,
        "isTag": 0,
        "description": "每个用户平均访问的次数"
    },
    {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 3,
        "name": "访问用户数",
        "bizName": "uv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["UV", "访问人数"],
        "relatedSchemaElements": [{ "dimensionId": 1, "nece
        "defaultAgg": "count",
        "order": 0.004658766300494777,
        "isTag": 0,
        "description": "访问的用户个数"
    }
],
"dimensions": [],
"aggType": "NONE",
"filterType": "AND",
"dimensionFilters": [],
"metricFilters": [],
"orders": [],
"dateInfo": {
    "dateMode": "RECENT",
    "startDate": "2024-06-08",
    "endDate": "2024-07-07",
    "dateList": [
        "2024-06-08",
        "2024-06-09",

```

```
        "2024-06-10",
        "2024-06-11",
        "2024-06-12",
        "2024-06-13",
        "2024-06-14",
        "2024-06-15",
        "2024-06-16",
        "2024-06-17",
        "2024-06-18",
        "2024-06-19",
        "2024-06-20",
        "2024-06-21",
        "2024-06-22",
        "2024-06-23",
        "2024-06-24",
        "2024-06-25",
        "2024-06-26",
        "2024-06-27",
        "2024-06-28",
        "2024-06-29",
        "2024-06-30",
        "2024-07-01",
        "2024-07-02",
        "2024-07-03",
        "2024-07-04",
        "2024-07-05",
        "2024-07-06",
        "2024-07-07"
    ],
    "unit": 30,
    "period": "DAY",
    "detectWord": "近30天",
    "groupByDate": false,
    "inherited": true,
    "groupByTimeDimension": "sys_imp_date"
},
"limit": 365,
"score": 4.0,
"elementMatches": [
    {
```

```

    "element": {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "model": 2,
      "id": 1,
      "name": "访问次数",
      "bizName": "pv",
      "useCnt": 0,
      "type": "METRIC",
      "alias": [],
      "relatedSchemaElements": [
        { "dimensionId": 1, "necessary": false },
        { "dimensionId": 2, "necessary": false }
      ],
      "defaultAgg": "SUM",
      "order": 0.0,
      "isTag": 0,
      "description": "一段时间内用户的访问次数"
    },
    "similarity": 1.0,
    "detectWord": "访问次数",
    "word": "访问次数",
    "frequency": 100000,
    "inherited": false
  },
  {
    "element": {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "model": 2,
      "id": 4,
      "name": "人均访问次数",
      "bizName": "pv_avg",
      "useCnt": 0,
      "type": "METRIC",
      "alias": ["平均访问次数"],
      "relatedSchemaElements": [
        { "dimensionId": 1, "necessary": false }
      ],
      "defaultAgg": "",

```

```

        "order": 0.002841870251016343,
        "isTag": 0,
        "description": "每个用户平均访问的次数"
    },
    "similarity": 0.9971581297489837,
    "detectWord": "人访问次数",
    "word": "人均访问次数",
    "frequency": 100000,
    "inherited": false
},
{
    "element": {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 3,
        "name": "访问用户数",
        "bizName": "uv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["UV", "访问人数"],
        "relatedSchemaElements": [
            { "dimensionId": 1, "necessary": false }
        ],
        "defaultAgg": "count",
        "order": 0.004658766300494777,
        "isTag": 0,
        "description": "访问的用户个数"
    },
    "similarity": 0.9953412336995052,
    "detectWord": "访问次数",
    "word": "访问用户数",
    "frequency": 100000,
    "inherited": false
}
],
"properties": {},
"sqlInfo": {},
"sqlEvaluation": {},
"queryType": "ID",

```



```

        "dataSetId": 1
    },
    "queryMode": "METRIC_MODEL"
}
],
"mapInfo": {
    "dataSetElementMatches": {
        "1": [
            {
                "element": {
                    "dataSet": 1,
                    "dataSetName": "超音数数据集",
                    "model": 2,
                    "id": 1,
                    "name": "访问次数",
                    "bizName": "pv",
                    "useCnt": 0,
                    "type": "METRIC",
                    "alias": [],
                    "relatedSchemaElements": [
                        { "dimensionId": 1, "necessary": false },
                        { "dimensionId": 2, "necessary": false }
                    ],
                    "defaultAgg": "SUM",
                    "order": 0.0,
                    "isTag": 0,
                    "description": "一段时间内用户的访问次数"
                },
                "similarity": 1.0,
                "detectWord": "访问次数",
                "word": "访问次数",
                "frequency": 100000,
                "inherited": false
            },
            {
                "element": {
                    "dataSet": 1,
                    "dataSetName": "超音数数据集",
                    "model": 2,
                    "id": 4,

```

```

        "name": "人均访问次数",
        "bizName": "pv_avg",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["平均访问次数"],
        "relatedSchemaElements": [{ "dimensionId": 1, "nece
        "defaultAgg": "",
        "order": 0.002841870251016343,
        "isTag": 0,
        "description": "每个用户平均访问的次数"
    },
    "similarity": 0.9971581297489837,
    "detectWord": "人访问次数",
    "word": "人均访问次数",
    "frequency": 100000,
    "inherited": false
},
{
    "element": {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 3,
        "name": "访问用户数",
        "bizName": "uv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["UV", "访问人数"],
        "relatedSchemaElements": [{ "dimensionId": 1, "nece
        "defaultAgg": "count",
        "order": 0.004658766300494777,
        "isTag": 0,
        "description": "访问的用户个数"
    },
    "similarity": 0.9953412336995052,
    "detectWord": "访问次数",
    "word": "访问用户数",
    "frequency": 100000,
    "inherited": false
},

```

```

    {
      "element": {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 1,
        "id": 1,
        "name": "部门",
        "bizName": "department",
        "useCnt": 0,
        "type": "DIMENSION",
        "alias": [],
        "schemaValueMaps": [],
        "order": 0.0,
        "isTag": 1,
        "description": ""
      },
      "similarity": 1.0,
      "detectWord": "部门",
      "word": "部门",
      "frequency": 100000,
      "inherited": true
    }
  ]
},
"matchedDataSetInfos": [1]
},
"mapModeEnum": "STRICT",
"queryDataType": "ALL",
"llmConfig": {
  "provider": "OPEN_AI",
  "baseUrl": "https://api.openai.com/v1",
  "apiKey": "kGA4i6jY7QqG5xH+S0N++SzVWVKyAHes0q7NpoFVI7pgU14e",
  "modelName": "gpt-3.5-turbo",
  "temperature": 0.0,
  "timeOut": 60
},
"exemplars": []
}

```

解释：

顶层结构

```
{
  "queryText": "超音速对比所有人访问次数",
  "chatId": 1,
  "dataSetIds": [],
  "modelIdToDataSetIds": { "1": [1], "2": [1], "3": [1], "4":
[2] },
  "user": {...},
  "saveAnswer": true,
  "text2SQLType": "RULE_AND_LLM",
  "candidateQueries": [...],
  "mapInfo": {...},
  "mapModeEnum": "STRICT",
  "queryDataType": "ALL",
  "llmConfig": {...},
  "exemplars": []
}
```

1. **queryText** : "超音速对比所有人访问次数"
 - 用户输入的查询文本，用户希望查询“超音速对比所有人访问次数”。
2. **chatId** : 1
 - 对话的唯一标识符，表示这个查询属于的对话ID。
3. **dataSetIds** : []
 - 空数组，通常用于指定查询涉及的数据集ID。这里为空，可能意味着在这次查询中并未显式指定数据集。
4. **modelIdToDataSetIds** : { "1": [1], "2": [1], "3": [1], "4": [2] }
 - 模型ID到数据集ID的映射关系。例如，模型ID为1、2、3的都对应数据集ID为1，而模型ID为4的对应数据集ID为2。
5. **user** :

```
{
  "id": 6,
  "name": "admin",
```

```
"displayName": "admin",
"email": "null",
"isAdmin": 1,
"superAdmin": true}
```

- 用户信息，包括用户ID、用户名、显示名称、邮箱、管理员权限和超级管理员标识。

6. `saveAnswer` : `true`

- 表示查询结果是否需要保存，`true` 表示结果将被保存。

7. `text2SQLType` : `"RULE_AND_LLM"`

- 使用的自然语言到SQL的转换策略类型，`RULE_AND_LLM` 表示结合了规则和大语言模型（LLM）的方法。

8. `candidateQueries` :

- 包含多个候选查询，每个候选查询是系统生成的可能的SQL查询版本。

9. `mapInfo` :

- 包含数据集元素匹配信息和匹配的数据集信息，详细解释见下文。

10. `mapModeEnum` : `"STRICT"`

- 映射模式，`STRICT` 表示严格模式，可能对查询解析的精确性要求较高。

11. `queryDataType` : `"ALL"`

- 查询的数据类型，`ALL` 表示查询包含所有类型的数据。

12. `llmConfig` :

- 大语言模型（LLM）相关的配置信息，详细解释见下文。

13. `exemplars` : `[]`

- 空数组，通常用于存储查询示例或训练数据。在此例中没有提供示例。

`candidateQueries` 数组中的项

候选查询是系统基于输入生成的多种可能的SQL查询版本，系统会从中选择一个或多个进行执行。

第一个候选查询

```
{
  "parseInfo": {
```

```

    "queryMode": "LLM_S2SQL",
    "dataSet": {...},
    "metrics": [],
    "dimensions": [],
    "aggType": "NONE",
    "filterType": "AND",
    "dimensionFilters": [],
    "metricFilters": [],
    "orders": [],
    "score": 24.0,
    "elementMatches": [...],
    "properties": {...},
    "sqlInfo": {...},
    "sqlEvaluation": {},
    "queryType": "ID",
    "dataSetId": 1
  },
  "queryMode": "LLM_S2SQL"
}

```

1. `parseInfo.queryMode` : "LLM_S2SQL"
 - 查询模式为使用大语言模型生成SQL (LLM_S2SQL)。
2. `parseInfo.dataSet` :
 - 与查询相关的数据集信息。
3. `parseInfo.metrics` : []
 - 空数组，表示此查询未涉及具体的度量指标。
4. `parseInfo.dimensions` : []
 - 空数组，表示此查询未涉及具体的维度。
5. `parseInfo.aggType` : "NONE"
 - 查询没有进行额外的聚合操作。
6. `parseInfo.filterType` : "AND"
 - 查询过滤条件使用的是 AND 逻辑。
7. `parseInfo.dimensionFilters` : []
 - 空数组，表示未使用维度过滤条件。

8. `parseInfo.metricFilters` : `[]`
 - 空数组，表示未使用指标过滤条件。
9. `parseInfo.orders` : `[]`
 - 空数组，表示没有指定查询结果的排序顺序。
10. `parseInfo.score` : `24.0`
 - 系统对该解析结果的评分，分数越高表示该解析结果与原始查询的匹配度越高。
11. `parseInfo.elementMatches` :
 - 查询中识别出的元素匹配信息。
12. `parseInfo.properties` :
 - 额外的属性信息。
13. `parseInfo.sqlInfo` :
 - SQL生成的相关信息。
14. `parseInfo.sqlEvaluation` : `{}`
 - 空对象，表示未进行SQL评估。
15. `parseInfo.queryType` : `"ID"`
 - 查询的类型，这里是ID类型查询，可能与系统内部逻辑相关。
16. `parseInfo.dataSetId` : `1`
 - 使用的数据集ID为1。
17. `queryMode` : `"LLM_S2SQL"`
 - 与 `parseInfo.queryMode` 相同，表示使用大语言模型生成SQL的查询模式。

第二个候选查询

```
{
  "parseInfo": {
    "queryMode": "METRIC_GROUPBY",
    "dataSet": {...},
    "metrics": [...],
    "dimensions": [...],
    "aggType": "NONE",
    "filterType": "AND",
    "dimensionFilters": [],
```

```

    "metricFilters": [],
    "orders": [],
    "dateInfo": {...},
    "limit": 365,
    "score": 6.0,
    "elementMatches": [...],
    "properties": {},
    "sqlInfo": {},
    "sqlEvaluation": {},
    "queryType": "ID",
    "dataSetId": 1
  },
  "queryMode": "METRIC_GROUPBY"
}

```

1. `parseInfo.queryMode` : "METRIC_GROUPBY"
 - 查询模式为按维度分组计算度量 (METRIC_GROUPBY)。
2. `parseInfo.dataSet` :
 - 与查询相关的数据集信息。
3. `parseInfo.metrics` :
 - 包含查询涉及的度量指标信息。
4. `parseInfo.dimensions` :
 - 包含查询涉及的维度信息。
5. `parseInfo.aggType` : "NONE"
 - 查询没有进行额外的聚合操作。
6. `parseInfo.filterType` : "AND"
 - 查询过滤条件使用的是 AND 逻辑。
7. `parseInfo.dimensionFilters` : []
 - 空数组，表示未使用维度过滤条件。
8. `parseInfo.metricFilters` : []
 - 空数组，表示未使用指标过滤条件。
9. `parseInfo.orders` : []
 - 空数组，表示没有指定查询结果的排序顺序。

10. `parseInfo.dateInfo` :
 - 查询涉及的时间范围信息。
11. `parseInfo.limit` : `365`
 - 查询结果的数量限制为365条记录。
12. `parseInfo.score` : `6.0`
 - 系统对该解析结果的评分，分数较低表示该解析结果与原始查询的匹配度较低。
13. `parseInfo.elementMatches` :
 - 查询中识别出的元素匹配信息。
14. `parseInfo.properties` : `{}`
 - 空对象，表示没有额外的属性信息。
15. `parseInfo.sqlInfo` : `{}`
 - 空对象，表示还没有生成SQL查询语句。
16. `parseInfo.sqlEvaluation` : `{}`
 - 空对象，表示没有进行SQL评估。
17. `parseInfo.queryType` : `"ID"`
 - 查询的类型，这里是ID类型查询，可能与系统内部逻辑相关。
18. `parseInfo.dataSetId` : `1`
 - 使用的数据集ID为1。
19. `queryMode` : `"METRIC_GROUPBY"`
 - 与 `parseInfo.queryMode` 相同，表示查询模式为按维度分组计算度量。

第三个候选查询

```
{
  "parseInfo": {
    "queryMode": "METRIC_MODEL",
    "dataSet": {...},
    "metrics": [...],
    "dimensions": [],
    "aggType": "NONE",
    "filterType": "AND",
    "dimensionFilters": [],
```

```

    "metricFilters": [],
    "orders": [],
    "dateInfo": {...},
    "limit": 365,
    "score": 4.0,
    "elementMatches": [...],
    "properties": {},
    "sqlInfo": {},
    "sqlEvaluation": {},
    "queryType": "ID",
    "dataSetId": 1
  },
  "queryMode": "METRIC_MODEL"
}

```

1. `parseInfo.queryMode` : "METRIC_MODEL"
 - 查询模式为度量模型 (METRIC_MODEL)。
2. `parseInfo.dataSet` :
 - 与查询相关的数据集信息。
3. `parseInfo.metrics` :
 - 包含查询涉及的度量指标信息。
4. `parseInfo.dimensions` : []
 - 空数组，表示此查询未涉及具体的维度。
5. `parseInfo.aggType` : "NONE"
 - 查询没有进行额外的聚合操作。
6. `parseInfo.filterType` : "AND"
 - 查询过滤条件使用的是 AND 逻辑。
7. `parseInfo.dimensionFilters` : []
 - 空数组，表示未使用维度过滤条件。
8. `parseInfo.metricFilters` : []
 - 空数组，表示未使用指标过滤条件。
9. `parseInfo.orders` : []
 - 空数组，表示没有指定查询结果的排序顺序。

10. `parseInfo.dateInfo` :
 - 查询涉及的时间范围信息。
11. `parseInfo.limit` : `365`
 - 查询结果的数量限制为365条记录。
12. `parseInfo.score` : `4.0`
 - 系统对该解析结果的评分，分数较低表示该解析结果与原始查询的匹配度较低。
13. `parseInfo.elementMatches` :
 - 查询中识别出的元素匹配信息。
14. `parseInfo.properties` : `{}`
 - 空对象，表示没有额外的属性信息。
15. `parseInfo.sqlInfo` : `{}`
 - 空对象，表示还没有生成SQL查询语句。
16. `parseInfo.sqlEvaluation` : `{}`
 - 空对象，表示没有进行SQL评估。
17. `parseInfo.queryType` : `"ID"`
 - 查询的类型，这里是ID类型查询，可能与系统内部逻辑相关。
18. `parseInfo.dataSetId` : `1`
 - 使用的数据集ID为1。
19. `queryMode` : `"METRIC_MODEL"`
 - 与 `parseInfo.queryMode` 相同，表示查询模式为度量模型。

`mapInfo` 对象

```
{
  "dataSetElementMatches": {
    "1": [
      {...},
      {...},
      {...},
      {...}
    ]
  },
}
```

```
"matchedDataSetInfos": [1]
}
```

1. **dataSetElementMatches** :
 - 这是一个映射，键是数据集ID，值是该数据集中匹配的元素列表。
2. **matchedDataSetInfos** : [1]
 - 这是一个数组，列出了匹配的数据集ID。在这个例子中，只有数据集ID为1的被匹配。

llmConfig 对象

```
{
  "provider": "OPEN_AI",
  "baseUrl": "https://api.openai.com/v1",
  "apiKey": "",
  "modelName": "gpt-3.5-turbo",
  "temperature": 0.0,
  "timeOut": 60
}
```

1. **provider** : "OPEN_AI"
 - 使用的LLM提供商是OpenAI。
2. **baseUrl** : "https://api.openai.com/v1"
 - LLM API的基础URL。
3. **apiKey** :
 - API调用所需的密钥。
4. **modelName** : "gpt-3.5-turbo"
 - 使用的模型名称，表示使用了OpenAI的GPT-3.5 Turbo模型。
5. **temperature** : 0.0
 - 生成文本的“温度”参数，控制输出的随机性。0.0 表示输出结果更加确定性，较少随机性。
6. **timeOut** : 60
 - API请求的超时时间，单位为秒。

关键点总结

1. **查询模式**：这个JSON定义了三个候选查询模式：`LLM_S2SQL`、`METRIC_GROUPBY` 和 `METRIC_MODEL`，分别处理不同的查询逻辑。
2. **用户信息**：请求由用户 `admin` 发起，且为管理员身份。
3. **数据集匹配**：查询的文本与数据集中的指标和维度进行了匹配，并选择了合适的模式生成候选SQL查询。
4. **LLM配置**：使用了OpenAI的GPT-3.5 Turbo模型，结合规则和LLM来处理自然语言查询并生成SQL。

14:57:19 [http-nio-9080-exec-2] DEBUG
c.t.s.h.s.w.s.i.WorkflowServiceImpl 67 -
QueryTypeParser result:

```
{
  "queryText": "超音速对比所有人访问次数",
  "chatId": 1,
  "dataSetIds": [],
  "modelIdToDataSetIds": { "1": [1], "2": [1], "3": [1], "4": [
  "user": {
    "id": 6,
    "name": "admin",
    "displayName": "admin",
    "email": "null",
    "isAdmin": 1,
    "superAdmin": true
  },
  "saveAnswer": true,
  "text2SQLType": "RULE_AND_LLM",
  "candidateQueries": [
    {
      "parseInfo": {
        "queryMode": "LLM_S2SQL",
        "dataSet": {
          "dataSet": 1,
          "dataSetName": "超音数数据集",
          "id": 1,
```

```

    "name": "超音数数据集",
    "bizName": "s2",
    "type": "DATASET",
    "order": 0.0,
    "isTag": 0
  },
  "metrics": [],
  "dimensions": [],
  "aggType": "NONE",
  "filterType": "AND",
  "dimensionFilters": [],
  "metricFilters": [],
  "orders": [],
  "score": 24.0,
  "elementMatches": [
    {
      "element": {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 1,
        "name": "访问次数",
        "bizName": "pv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": [],
        "relatedSchemaElements": [
          { "dimensionId": 1, "necessary": false },
          { "dimensionId": 2, "necessary": false }
        ],
        "defaultAgg": "SUM",
        "order": 0.0,
        "isTag": 0,
        "description": "一段时间内用户的访问次数"
      },
      "similarity": 1.0,
      "detectWord": "访问次数",
      "word": "访问次数",
      "frequency": 100000,
      "inherited": false
    }
  ]
}

```

```

    },
    {
      "element": {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 4,
        "name": "人均访问次数",
        "bizName": "pv_avg",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["平均访问次数"],
        "relatedSchemaElements": [
          { "dimensionId": 1, "necessary": false }
        ],
        "defaultAgg": "",
        "order": 0.002841870251016343,
        "isTag": 0,
        "description": "每个用户平均访问的次数"
      },
      "similarity": 0.9971581297489837,
      "detectWord": "人访问次数",
      "word": "人均访问次数",
      "frequency": 100000,
      "inherited": false
    },
    {
      "element": {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 3,
        "name": "访问用户数",
        "bizName": "uv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["UV", "访问人数"],
        "relatedSchemaElements": [
          { "dimensionId": 1, "necessary": false }
        ],

```

```

        "defaultAgg": "count",
        "order": 0.004658766300494777,
        "isTag": 0,
        "description": "访问的用户个数"
    },
    "similarity": 0.9953412336995052,
    "detectWord": "访问次数",
    "word": "访问用户数",
    "frequency": 100000,
    "inherited": false
},
{
    "element": {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 1,
        "id": 1,
        "name": "部门",
        "bizName": "department",
        "useCnt": 0,
        "type": "DIMENSION",
        "alias": [],
        "schemaValueMaps": [],
        "order": 0.0,
        "isTag": 1,
        "description": ""
    },
    "similarity": 1.0,
    "detectWord": "部门",
    "word": "部门",
    "frequency": 100000,
    "inherited": true
}
],
"properties": {
    "type": "internal",
    "CONTEXT": {
        "dataSetId": 1,
        "llmReq": {
            "queryText": "超音速对比所有人访问次数",

```



```
"filterCondition": {},
"schema": {
  "domainName": "超音数数据集",
  "dataSetName": "超音数数据集",
  "dataSetId": 1,
  "fieldNameList": [
    "访问用户数",
    "用户",
    "部门",
    "访问次数",
    "人均访问次数",
    "停留时长",
    "页面",
    "数据日期"
  ],
  "metrics": [
    {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "model": 2,
      "id": 1,
      "name": "访问次数",
      "bizName": "pv",
      "useCnt": 0,
      "type": "METRIC",
      "alias": [],
      "relatedSchemaElements": [
        { "dimensionId": 1, "necessary": false },
        { "dimensionId": 2, "necessary": false }
      ],
      "defaultAgg": "SUM",
      "order": 0.0,
      "isTag": 0,
      "description": "一段时间内用户的访问次数"
    },
    {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "model": 2,
      "id": 4,
```

```

        "name": "人均访问次数",
        "bizName": "pv_avg",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["平均访问次数"],
        "relatedSchemaElements": [
            { "dimensionId": 1, "necessary": false }
        ],
        "defaultAgg": "",
        "order": 0.002841870251016343,
        "isTag": 0,
        "description": "每个用户平均访问的次数"
    },
    {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 3,
        "name": "访问用户数",
        "bizName": "uv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["UV", "访问人数"],
        "relatedSchemaElements": [
            { "dimensionId": 1, "necessary": false }
        ],
        "defaultAgg": "count",
        "order": 0.004658766300494777,
        "isTag": 0,
        "description": "访问的用户个数"
    }
],
"dimensions": [
    {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 1,
        "id": 1,
        "name": "部门",
        "bizName": "department",

```

```

        "useCnt": 0,
        "type": "DIMENSION",
        "alias": [],
        "schemaValueMaps": [],
        "order": 0.0,
        "isTag": 1,
        "description": ""
    }
],
    "terms": []
},
    "linking": [],
    "currentDate": "2024-08-07",
    "priorExts": "",
    "sqlGenType": "1_pass_self_consistency",
    "llmConfig": {
        "provider": "OPEN_AI",
        "baseUrl": "https://api.openai.com/v1",
        "apiKey": "kGA4i6jY7QqG5xH+S0N++SzVWVKyAHes0q7N",
        "modelName": "gpt-3.5-turbo",
        "temperature": 0.0,
        "timeOut": 60
    },
    "exemplars": []
},
    "llmResp": {
        "query": "超音速对比所有人访问次数",
        "modelName": "超音数数据集",
        "dbSchema": "Table: 超音数数据集, Metrics: [访问次数",
        "sqlOutput": "SELECT SUM(访问次数) FROM 超音数数据集",
        "sqlRespMap": {
            "SELECT SUM(访问次数) FROM 超音数数据集": {
                "sqlWeight": 1.0,
                "fewShots": [
                    {
                        "question": "今年以来发行的歌曲中, 有哪些在近7",
                        "dbSchema": "Table: 歌曲库, Columns = [\"发",
                        "sql": "SELECT 歌曲名 FROM 歌曲库 WHERE dat
                    },
                    {

```

```

        "question": "超音数近半年每个月的平均访问次数",
        "dbSchema": "Table: 超音数产品, Columns = [",
        "sql": "SELECT MONTH(数据日期), AVG(访问次数",
    },
    {
        "question": "我想要近半年签约的播放量前十的歌手",
        "dbSchema": "Table: 艺人库, Columns = [\"播",
        "sql": "SELECT 歌手名 FROM 艺人库 WHERE dat",
    },
    {
        "question": "播放量大于1万的歌曲有多少 (补充信",
        "dbSchema": "Table: 歌曲库, Columns = [\"歌",
        "sql": "SELECT 歌曲名 FROM 歌曲库 WHERE 结算",
    },
    {
        "question": "超音数访问次数大于1k的部门是哪些",
        "dbSchema": "Table: 超音数产品, Columns = [",
        "sql": "SELECT 部门 FROM 超音数产品 WHERE 访",
    }
]
}
}
},
"linkingValues": []
}
},
"sqlInfo": {
    "s2SQL": "SELECT SUM(访问次数) FROM 超音数数据集",
    "correctS2SQL": "SELECT SUM(访问次数) FROM 超音数数据集",
},
"sqlEvaluation": {},
"queryType": "METRIC",
"datasetId": 1
},
"queryMode": "LLM_S2SQL"
},
{
    "parseInfo": {
        "queryMode": "METRIC_GROUPBY",
        "dataset": {

```

```

    "dataSet": 1,
    "dataSetName": "超音数数据集",
    "id": 1,
    "name": "超音数数据集",
    "bizName": "s2",
    "type": "DATASET",
    "order": 0.0,
    "isTag": 0
  },
  "metrics": [
    {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "model": 2,
      "id": 1,
      "name": "访问次数",
      "bizName": "pv",
      "useCnt": 0,
      "type": "METRIC",
      "alias": [],
      "relatedSchemaElements": [
        { "dimensionId": 1, "necessary": false },
        { "dimensionId": 2, "necessary": false }
      ],
      "defaultAgg": "SUM",
      "order": 0.0,
      "isTag": 0,
      "description": "一段时间内用户的访问次数"
    },
    {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "model": 2,
      "id": 4,
      "name": "人均访问次数",
      "bizName": "pv_avg",
      "useCnt": 0,
      "type": "METRIC",
      "alias": ["平均访问次数"],
      "relatedSchemaElements": [{ "dimensionId": 1, "nece

```

```

        "defaultAgg": "",
        "order": 0.002841870251016343,
        "isTag": 0,
        "description": "每个用户平均访问的次数"
    },
    {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 3,
        "name": "访问用户数",
        "bizName": "uv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["UV", "访问人数"],
        "relatedSchemaElements": [{ "dimensionId": 1, "nece
        "defaultAgg": "count",
        "order": 0.004658766300494777,
        "isTag": 0,
        "description": "访问的用户个数"
    }
],
"dimensions": [
    { "bizName": "sys_imp_date", "order": 0.0, "isTag": 0
    {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 1,
        "id": 1,
        "name": "部门",
        "bizName": "department",
        "useCnt": 0,
        "type": "DIMENSION",
        "alias": [],
        "schemaValueMaps": [],
        "order": 0.0,
        "isTag": 1,
        "description": ""
    }
],

```

```
"aggType": "NONE",
"filterType": "AND",
"dimensionFilters": [],
"metricFilters": [],
"orders": [],
"dateInfo": {
  "dateMode": "RECENT",
  "startDate": "2024-06-08",
  "endDate": "2024-07-07",
  "dateList": [
    "2024-06-08",
    "2024-06-09",
    "2024-06-10",
    "2024-06-11",
    "2024-06-12",
    "2024-06-13",
    "2024-06-14",
    "2024-06-15",
    "2024-06-16",
    "2024-06-17",
    "2024-06-18",
    "2024-06-19",
    "2024-06-20",
    "2024-06-21",
    "2024-06-22",
    "2024-06-23",
    "2024-06-24",
    "2024-06-25",
    "2024-06-26",
    "2024-06-27",
    "2024-06-28",
    "2024-06-29",
    "2024-06-30",
    "2024-07-01",
    "2024-07-02",
    "2024-07-03",
    "2024-07-04",
    "2024-07-05",
    "2024-07-06",
    "2024-07-07"
```

```

    ],
    "unit": 30,
    "period": "DAY",
    "detectWord": "近30天",
    "groupByDate": false,
    "inherited": true,
    "groupByTimeDimension": "sys_imp_date"
  },
  "limit": 365,
  "score": 6.0,
  "elementMatches": [
    {
      "element": {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 1,
        "name": "访问次数",
        "bizName": "pv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": [],
        "relatedSchemaElements": [
          { "dimensionId": 1, "necessary": false },
          { "dimensionId": 2, "necessary": false }
        ],
        "defaultAgg": "SUM",
        "order": 0.0,
        "isTag": 0,
        "description": "一段时间内用户的访问次数"
      },
      "similarity": 1.0,
      "detectWord": "访问次数",
      "word": "访问次数",
      "frequency": 100000,
      "inherited": false
    },
    {
      "element": {
        "dataSet": 1,

```



```

    "dataSetName": "超音数数据集",
    "model": 2,
    "id": 4,
    "name": "人均访问次数",
    "bizName": "pv_avg",
    "useCnt": 0,
    "type": "METRIC",
    "alias": ["平均访问次数"],
    "relatedSchemaElements": [
      { "dimensionId": 1, "necessary": false }
    ],
    "defaultAgg": "",
    "order": 0.002841870251016343,
    "isTag": 0,
    "description": "每个用户平均访问的次数"
  },
  "similarity": 0.9971581297489837,
  "detectWord": "人访问次数",
  "word": "人均访问次数",
  "frequency": 100000,
  "inherited": false
},
{
  "element": {
    "dataSet": 1,
    "dataSetName": "超音数数据集",
    "model": 2,
    "id": 3,
    "name": "访问用户数",
    "bizName": "uv",
    "useCnt": 0,
    "type": "METRIC",
    "alias": ["UV", "访问人数"],
    "relatedSchemaElements": [
      { "dimensionId": 1, "necessary": false }
    ],
    "defaultAgg": "count",
    "order": 0.004658766300494777,
    "isTag": 0,
    "description": "访问的用户个数"
  }
}

```

```

    },
    "similarity": 0.9953412336995052,
    "detectWord": "访问次数",
    "word": "访问用户数",
    "frequency": 100000,
    "inherited": false
  },
  {
    "element": {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "model": 1,
      "id": 1,
      "name": "部门",
      "bizName": "department",
      "useCnt": 0,
      "type": "DIMENSION",
      "alias": [],
      "schemaValueMaps": [],
      "order": 0.0,
      "isTag": 1,
      "description": ""
    },
    "similarity": 1.0,
    "detectWord": "部门",
    "word": "部门",
    "frequency": 100000,
    "inherited": true
  }
],
"properties": {},
"sqlInfo": {
  "s2SQL": "SELECT sys_imp_date, 部门, SUM(访问次数) AS 访",
  "correctS2SQL": "SELECT sys_imp_date, 部门, SUM(访问次数"
},
"sqlEvaluation": {},
"queryType": "METRIC",
"dataSetId": 1
},
"queryMode": "METRIC_GROUPBY"

```

```

},
{
  "parseInfo": {
    "queryMode": "METRIC_MODEL",
    "dataSet": {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "id": 1,
      "name": "超音数数据集",
      "bizName": "s2",
      "type": "DATASET",
      "order": 0.0,
      "isTag": 0
    },
    "metrics": [
      {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 1,
        "name": "访问次数",
        "bizName": "pv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": [],
        "relatedSchemaElements": [
          { "dimensionId": 1, "necessary": false },
          { "dimensionId": 2, "necessary": false }
        ],
        "defaultAgg": "SUM",
        "order": 0.0,
        "isTag": 0,
        "description": "一段时间内用户的访问次数"
      },
      {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 4,
        "name": "人均访问次数",

```

```

        "bizName": "pv_avg",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["平均访问次数"],
        "relatedSchemaElements": [{ "dimensionId": 1, "nece
        "defaultAgg": "",
        "order": 0.002841870251016343,
        "isTag": 0,
        "description": "每个用户平均访问的次数"
    },
    {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 3,
        "name": "访问用户数",
        "bizName": "uv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["UV", "访问人数"],
        "relatedSchemaElements": [{ "dimensionId": 1, "nece
        "defaultAgg": "count",
        "order": 0.004658766300494777,
        "isTag": 0,
        "description": "访问的用户个数"
    }
],
"dimensions": [{ "bizName": "sys_imp_date", "order": 0.
"aggType": "NONE",
"filterType": "AND",
"dimensionFilters": [],
"metricFilters": [],
"orders": [],
"dateInfo": {
    "dateMode": "RECENT",
    "startDate": "2024-06-08",
    "endDate": "2024-07-07",
    "dateList": [
        "2024-06-08",
        "2024-06-09",

```

```
        "2024-06-10",
        "2024-06-11",
        "2024-06-12",
        "2024-06-13",
        "2024-06-14",
        "2024-06-15",
        "2024-06-16",
        "2024-06-17",
        "2024-06-18",
        "2024-06-19",
        "2024-06-20",
        "2024-06-21",
        "2024-06-22",
        "2024-06-23",
        "2024-06-24",
        "2024-06-25",
        "2024-06-26",
        "2024-06-27",
        "2024-06-28",
        "2024-06-29",
        "2024-06-30",
        "2024-07-01",
        "2024-07-02",
        "2024-07-03",
        "2024-07-04",
        "2024-07-05",
        "2024-07-06",
        "2024-07-07"
    ],
    "unit": 30,
    "period": "DAY",
    "detectWord": "近30天",
    "groupByDate": false,
    "inherited": true,
    "groupByTimeDimension": "sys_imp_date"
},
"limit": 365,
"score": 4.0,
"elementMatches": [
    {
```

```

    "element": {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "model": 2,
      "id": 1,
      "name": "访问次数",
      "bizName": "pv",
      "useCnt": 0,
      "type": "METRIC",
      "alias": [],
      "relatedSchemaElements": [
        { "dimensionId": 1, "necessary": false },
        { "dimensionId": 2, "necessary": false }
      ],
      "defaultAgg": "SUM",
      "order": 0.0,
      "isTag": 0,
      "description": "一段时间内用户的访问次数"
    },
    "similarity": 1.0,
    "detectWord": "访问次数",
    "word": "访问次数",
    "frequency": 100000,
    "inherited": false
  },
  {
    "element": {
      "dataSet": 1,
      "dataSetName": "超音数数据集",
      "model": 2,
      "id": 4,
      "name": "人均访问次数",
      "bizName": "pv_avg",
      "useCnt": 0,
      "type": "METRIC",
      "alias": ["平均访问次数"],
      "relatedSchemaElements": [
        { "dimensionId": 1, "necessary": false }
      ],
      "defaultAgg": "",

```

```

        "order": 0.002841870251016343,
        "isTag": 0,
        "description": "每个用户平均访问的次数"
    },
    "similarity": 0.9971581297489837,
    "detectWord": "人访问次数",
    "word": "人均访问次数",
    "frequency": 100000,
    "inherited": false
},
{
    "element": {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 3,
        "name": "访问用户数",
        "bizName": "uv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["UV", "访问人数"],
        "relatedSchemaElements": [
            { "dimensionId": 1, "necessary": false }
        ],
        "defaultAgg": "count",
        "order": 0.004658766300494777,
        "isTag": 0,
        "description": "访问的用户个数"
    },
    "similarity": 0.9953412336995052,
    "detectWord": "访问次数",
    "word": "访问用户数",
    "frequency": 100000,
    "inherited": false
}
],
"properties": {},
"sqlInfo": {
    "s2SQL": "SELECT sys_imp_date, SUM(访问次数) AS 访问次数",
    "correctS2SQL": "SELECT sys_imp_date, SUM(访问次数) AS

```

```

    },
    "sqlEvaluation": {},
    "queryType": "METRIC",
    "dataSetId": 1
  },
  "queryMode": "METRIC_MODEL"
},
],
"mapInfo": {
  "dataSetElementMatches": {
    "1": [
      {
        "element": {
          "dataSet": 1,
          "dataSetName": "超音数数据集",
          "model": 2,
          "id": 1,
          "name": "访问次数",
          "bizName": "pv",
          "useCnt": 0,
          "type": "METRIC",
          "alias": [],
          "relatedSchemaElements": [
            { "dimensionId": 1, "necessary": false },
            { "dimensionId": 2, "necessary": false }
          ],
          "defaultAgg": "SUM",
          "order": 0.0,
          "isTag": 0,
          "description": "一段时间内用户的访问次数"
        },
        "similarity": 1.0,
        "detectWord": "访问次数",
        "word": "访问次数",
        "frequency": 100000,
        "inherited": false
      },
      {
        "element": {
          "dataSet": 1,

```



```

        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 4,
        "name": "人均访问次数",
        "bizName": "pv_avg",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["平均访问次数"],
        "relatedSchemaElements": [{ "dimensionId": 1, "nece
        "defaultAgg": "",
        "order": 0.002841870251016343,
        "isTag": 0,
        "description": "每个用户平均访问的次数"
    },
    "similarity": 0.9971581297489837,
    "detectWord": "人访问次数",
    "word": "人均访问次数",
    "frequency": 100000,
    "inherited": false
},
{
    "element": {
        "dataSet": 1,
        "dataSetName": "超音数数据集",
        "model": 2,
        "id": 3,
        "name": "访问用户数",
        "bizName": "uv",
        "useCnt": 0,
        "type": "METRIC",
        "alias": ["UV", "访问人数"],
        "relatedSchemaElements": [{ "dimensionId": 1, "nece
        "defaultAgg": "count",
        "order": 0.004658766300494777,
        "isTag": 0,
        "description": "访问的用户个数"
    },
    "similarity": 0.9953412336995052,
    "detectWord": "访问次数",
    "word": "访问用户数",

```

```

        "frequency": 100000,
        "inherited": false
    },
    {
        "element": {
            "dataSet": 1,
            "dataSetName": "超音数数据集",
            "model": 1,
            "id": 1,
            "name": "部门",
            "bizName": "department",
            "useCnt": 0,
            "type": "DIMENSION",
            "alias": [],
            "schemaValueMaps": [],
            "order": 0.0,
            "isTag": 1,
            "description": ""
        },
        "similarity": 1.0,
        "detectWord": "部门",
        "word": "部门",
        "frequency": 100000,
        "inherited": true
    }
]
},
"matchedDataSetInfos": [1]
},
"mapModeEnum": "STRICT",
"queryDataType": "ALL",
"llmConfig": {
    "provider": "OPEN_AI",
    "baseUrl": "https://api.openai.com/v1",
    "apiKey": "kGA4i6jY7QqG5xH+S0N++SzVWVKyAHes0q7NpoFVI7pgU14e",
    "modelName": "gpt-3.5-turbo",
    "temperature": 0.0,
    "timeOut": 60
},

```

```
"exemplars": []
}
```

解释

顶层结构分析

```
{
  "queryText": "超音速对比所有人访问次数",
  "chatId": 1,
  "dataSetIds": [],
  "modelIdToDataSetIds": { "1": [1], "2": [1], "3": [1], "4":
[2] },
  "user": { ... },
  "saveAnswer": true,
  "text2SQLType": "RULE_AND_LLM",
  "candidateQueries": [ ... ],
  "mapInfo": { ... },
  "mapModeEnum": "STRICT",
  "queryDataType": "ALL",
  "llmConfig": { ... },
  "exemplars": []
}
```

1. **queryText** : "超音速对比所有人访问次数"
 - 用户的自然语言查询文本。用户希望查询“超音速”与“所有人”的访问次数进行对比。
2. **chatId** : 1
 - 这是对话的唯一标识符，表明该查询属于对话ID为1的会话。
3. **dataSetIds** : []
 - 空数组，表示没有显式指定数据集ID。查询的相关数据集将由系统自动选择或推断。
4. **modelIdToDataSetIds** : { "1": [1], "2": [1], "3": [1], "4": [2] }
 - 这是模型ID到数据集ID的映射。例如，模型ID为1、2、3的模型都映射到数据集ID为1，而模型ID为4的模型映射到数据集ID为2。
5. **user** :

- 用户信息的结构体，包含用户ID、用户名、显示名称、邮箱、管理员权限等信息。
6. `saveAnswer` : `true`
 - 表示查询结果是否应被保存。 `true` 表示结果将被保存。
 7. `text2SQLType` : `"RULE_AND_LLM"`
 - 指定了将自然语言转换为SQL的策略类型。 `RULE_AND_LLM` 表示结合了规则和大语言模型（LLM）的方法。
 8. `candidateQueries` :
 - 包含多个候选查询，每个候选查询是系统根据输入生成的不同SQL查询版本。
 9. `mapInfo` :
 - 包含了与数据集元素匹配的信息和查询的映射信息，详细解释见下文。
 10. `mapModeEnum` : `"STRICT"`
 - 映射模式， `STRICT` 表示严格模式，这意味着在生成SQL时会严格遵循匹配规则。
 11. `queryDataType` : `"ALL"`
 - 查询的数据类型为 `ALL`，表示查询涉及所有类型的数据。
 12. `llmConfig` :
 - 大语言模型（LLM）的相关配置，详细解释见下文。
 13. `exemplars` : `[]`
 - 空数组，通常用于存储查询示例或训练数据。在这个例子中没有提供示例。

用户信息 (`user`)

```
"user": {
  "id": 6,
  "name": "admin",
  "displayName": "admin",
  "email": "null",
  "isAdmin": 1,
  "superAdmin": true}
```

- `id` : `6`
 - 用户的唯一标识符。

- `name` : `"admin"`
 - 用户名。
- `displayName` : `"admin"`
 - 用户的显示名称。
- `email` : `"null"`
 - 用户的电子邮件地址，这里显示为 `null`，表示没有提供。
- `isAdmin` : `1`
 - 表示该用户是管理员，`1` 表示是管理员。
- `superAdmin` : `true`
 - 表示该用户是超级管理员，`true` 表示是超级管理员。

候选查询 (`candidateQueries`)

`candidateQueries` 包含多个候选查询，系统会从中选择最合适的执行。每个候选查询包含解析信息和SQL生成信息。

第一个候选查询

```
{
  "parseInfo": {
    "queryMode": "LLM_S2SQL",
    "dataSet": { ... },
    "metrics": [],
    "dimensions": [],
    "aggType": "NONE",
    "filterType": "AND",
    "dimensionFilters": [],
    "metricFilters": [],
    "orders": [],
    "score": 24.0,
    "elementMatches": [ ... ],
    "properties": { ... },
    "sqlInfo": { ... },
    "sqlEvaluation": {},
    "queryType": "METRIC",
    "dataSetId": 1
  },
}
```

```
"queryMode": "LLM_S2SQL"
}
```

1. **queryMode** : "LLM_S2SQL"
 - 表示这个候选查询是使用大语言模型（LLM）生成的SQL查询。
2. **dataset** :
 - 与查询相关的数据集信息，包含数据集的ID、名称等。
3. **metrics** : []
 - 空数组，表示这个候选查询没有涉及具体的度量指标。
4. **dimensions** : []
 - 空数组，表示这个候选查询没有涉及具体的维度。
5. **aggType** : "NONE"
 - 查询没有进行额外的聚合操作。
6. **filterType** : "AND"
 - 查询过滤条件使用的是 **AND** 逻辑。
7. **dimensionFilters** : []
 - 空数组，表示没有使用维度过滤条件。
8. **metricFilters** : []
 - 空数组，表示没有使用指标过滤条件。
9. **orders** : []
 - 空数组，表示没有指定查询结果的排序顺序。
10. **score** : 24.0
 - 系统对这个候选查询的评分，表示这个解析结果与原始查询的匹配度。
11. **elementMatches** :
 - 包含识别出的元素与数据集之间的匹配信息。
12. **properties** :
 - 额外的属性信息，例如上下文信息和 LLM 请求/响应信息。
13. **sqlInfo** :
 - 包含生成的SQL语句及其相关信息。

14. `sqlEvaluation`: `{}`
 - 空对象，表示没有进行SQL评估。
15. `queryType`: `"METRIC"`
 - 表示这个查询类型是度量查询。
16. `dataSetId`: `1`
 - 表示使用的数据集的ID是1。

第二个候选查询

```
{
  "parseInfo": {
    "queryMode": "METRIC_GROUPBY",
    "dataSet": { ... },
    "metrics": [ ... ],
    "dimensions": [ ... ],
    "aggType": "NONE",
    "filterType": "AND",
    "dimensionFilters": [],
    "metricFilters": [],
    "orders": [],
    "dateInfo": { ... },
    "limit": 365,
    "score": 6.0,
    "elementMatches": [ ... ],
    "properties": {},
    "sqlInfo": { ... },
    "sqlEvaluation": {},
    "queryType": "METRIC",
    "dataSetId": 1
  },
  "queryMode": "METRIC_GROUPBY"
}
```

1. `queryMode`: `"METRIC_GROUPBY"`
 - 表示这个候选查询是按维度分组的度量查询。
2. `metrics`:

- 包含这个查询涉及的度量指标的信息。
3. **dimensions** :
 - 包含这个查询涉及的维度信息。
 4. **dateInfo** :
 - 包含与查询相关的时间信息，如时间范围、分组方式等。
 5. **limit** : 365
 - 表示这个查询的结果数量限制为365条记录。
 6. **score** : 6.0
 - 这个候选查询的评分，分数较低表示与原始查询的匹配度相对较低。
 7. **elementMatches** :
 - 包含识别出的元素与数据集之间的匹配信息。
 8. **sqlInfo** :
 - 包含生成的SQL语句及其相关信息。
 9. **queryType** : "METRIC"
 - 表示这个查询类型是度量查询。
 10. **dataSetId** : 1
 - 表示使用的数据集的ID是1。

第三个候选查询

```
{
  "parseInfo": {
    "queryMode": "METRIC_MODEL",
    "dataSet": { ... },
    "metrics": [ ... ],
    "dimensions": [ ... ],
    "aggType": "NONE",
    "filterType": "AND",
    "dimensionFilters": [],
    "metricFilters": [],
    "orders": [],
    "dateInfo": { ... },
    "limit": 365,
```



```
    "score": 4.0,  
    "elementMatches": [ ... ],  
    "properties": {},  
    "sqlInfo": { ... },  
    "sqlEvaluation": {},  
    "queryType": "METRIC",  
    "dataSetId": 1  
  },  
  "queryMode": "METRIC_MODEL"  
}
```

1. **queryMode** : "METRIC_MODEL"
 - 表示这个候选查询是度量模型查询。
2. **metrics** :
 - 包含这个查询涉及的度量指标的信息。
3. **dimensions** :
 - 包含这个查询涉及的维度信息。
4. **dateInfo** :
 - 包含与查询相关的时间信息，如时间范围、分组方式等。
5. **limit** : 365
 - 表示这个查询的结果数量限制为365条记录。
6. **score** : 4.0
 - 这个候选查询的评分，分数较低表示与原始查询的匹配度相对较低。
7. **elementMatches** :
 - 包含识别出的元素与数据集之间的匹配信息。
8. **sqlInfo** :
 - 包含生成的SQL语句及其相关信息。
9. **queryType** : "METRIC"
 - 表示这个查询类型是度量查询。
10. **dataSetId** : 1
 - 表示使用的数据集的ID是1。

数据集映射信息 (`mapInfo`)

```
"mapInfo": {
  "dataSetElementMatches": {
    "1": [
      { ... },
      { ... },
      { ... },
      { ... }
    ]
  },
  "matchedDataSetInfos": [1]
}
```

1. `dataSetElementMatches` :

- 这是一个映射，键是数据集ID，值是该数据集中匹配的元素列表。这里表示数据集ID为1的匹配了多个元素。

2. `matchedDataSetInfos` : `[1]`

- 这是一个数组，列出了匹配的数据集ID。在这个例子中，只有数据集ID为1的被匹配。

大语言模型配置 (`llmConfig`)

```
"llmConfig": {
  "provider": "OPEN_AI",
  "baseUrl": "https://api.openai.com/v1",
  "apiKey": "",
  "modelName": "gpt-3.5-turbo",
  "temperature": 0.0,
  "timeOut": 60
}
```

1. `provider` : `"OPEN_AI"`

- 使用的LLM提供商是OpenAI。

2. `baseUrl` : `"https://api.openai.com/v1"`

- LLM API的基础URL。

3. `apiKey` :

- 用于API调用的密钥。
4. `modelName` : "gpt-3.5-turbo"
 - 使用的模型名称，表示使用了OpenAI的GPT-3.5 Turbo模型。
 5. `temperature` : 0.0
 - 生成文本的“温度”参数，控制输出的随机性。0.0 表示输出结果更加确定性，较少随机性。
 6. `timeOut` : 60
 - API请求的超时时间，单位为秒。

关键点总结

1. **查询模式**：JSON定义了三个候选查询模式，分别是 `LLM_S2SQL`、`METRIC_GROUPBY` 和 `METRIC_MODEL`，每种模式处理查询的方式不同。
2. **用户信息**：请求由用户 `admin` 发起，且该用户拥有管理员权限。
3. **数据集映射**：查询文本与数据集中的度量和维度进行了匹配，并选择了合适的模式生成SQL查询。
4. **LLM配置**：使用了OpenAI的GPT-3.5 Turbo模型，结合规则和LLM来处理自然语言查询并生成SQL。

总结

1. RuleSqlParser (规则解析器)

- **基本原理:**
 - `RuleSqlParser` 使用预定义的规则和模式将自然语言查询转换为 SQL 查询。这种方法通常依赖于手工编写的规则库，适用于已知和明确的查询模式。
- **使用场景:**
 - 适用于结构化且相对简单的查询。
 - 更加可靠和可控，因为规则是预先定义的。
 - 通常不涉及复杂的自然语言理解，而是基于模式匹配。
- **优点:**
 - 快速且高效，特别是在处理预定义查询时。
 - 可以避免复杂的自然语言处理问题，适用于较明确的查询。
- **缺点:**

- 灵活性较低，无法处理超出规则库范围的复杂查询。
- 依赖于预定义规则，无法应对新型或复杂的自然语言查询。

2. LLMSqlParser (大语言模型解析器)

- **基本原理:**
 - `LLMSqlParser` 利用大语言模型（LLM），如 GPT-3.5，来解析自然语言查询并生成 SQL 查询。LLM 能够理解更复杂的自然语言结构，并且能够处理灵活多变的查询表达方式。
- **使用场景:**
 - 适用于复杂、非结构化的自然语言查询。
 - 可以处理多样化的语言表达，适应性强。
- **优点:**
 - 更灵活，能够理解并生成复杂的 SQL 查询。
 - 适应性强，能够处理新颖或未定义的查询模式。
- **缺点:**
 - 生成的查询可能不如规则解析器那样可控和精准。
 - 依赖于大语言模型，可能会有更高的计算成本和时间消耗。

3. QueryTypeParser (查询类型解析器)

- **基本原理:**
 - `QueryTypeParser` 可能是一个根据查询类型（如度量查询、聚合查询等）进行分类的解析器。它根据查询的类型选择适当的解析策略来生成 SQL 查询。
- **使用场景:**
 - 根据查询的类型（例如度量查询、聚合查询、模型查询等）来选择最适合的解析方法。
 - 通常会结合其他解析器（如 `RuleSqlParser` 或 `LLMSqlParser`）进行最终的 SQL 生成。
- **优点:**
 - 能够自动选择最适合当前查询类型的解析器，从而提高解析效率和准确性。
 - 灵活性较强，适用于多种查询类型。
- **缺点:**

- 可能需要根据查询类型进行更多的前置判断和处理。

关键区别总结

- 解析方法: `RuleSqlParser` 使用规则和模板, `LLMSqlParser` 依赖于大语言模型, `QueryTypeParser` 根据查询类型选择合适的解析方法。