文章
姚 鑫 · 四月 25, 2021 阅读大约需 10 分钟

第五章 优化查询性能（四）

第五章 优化查询性能（四）

注释选项

可以在SELECT、INSERT、UPDATE、DELETE或TRUNCATE表命令中为查询优化器指定一个或多个注释选项。
注释选项指定查询优化器在编译SQL查询期间使用的选项。
通常，注释选项用于覆盖特定查询的系统范围默认配置。

语法

语法/*#OPTIONS */(在/*和#之间没有空格)指定了一个注释选项。
注释选项不是注释;
它为查询优化器指定一个值。
注释选项使用JSON语法指定，通常是 “ key:value ” 对，例如: /*#OPTIONS {"optionName":value} */。
支持更复杂的JSON语法，比如嵌套值。

注释选项不是注释;
除了JSON语法之外，它可能不包含任何文本。
包含非json文本在/* ... */分隔符导致SQLCODE -153错误。
InterSystems SQL不验证JSON字符串的内容。

#OPTIONS关键字必须用大写字母指定。
JSON的大括号语法中不应该使用空格。
如果SQL代码用引号括起来，比如动态SQL语句，JSON语法中的引号应该是双引号。
例如:myquery="SELECT Name FROM Sample.MyTest /*#OPTIONS {""optName"":""optValue""} */".

可以在SQL代码中任何可以指定注释的地方指定/*#OPTIONS */ comment选项。
在显示的语句文本中，注释选项总是作为注释显示在语句文本的末尾。

你可以在SQL代码中指定多个/*#OPTIONS */ comment选项。
它们按照指定的顺序显示在返回的语句文本中。
如果为同一个选项指定了多个注释选项，则使用last指定的选项值。

以下的注释选项被记录在案:

- /*#OPTIONS {"BiasAsOutlier":1} */
- /*#OPTIONS {"DynamicSQLTypeList":"10,1,11"}
- /*#OPTIONS {"NoTempFile":1} */

显示

/*#OPTIONS */ comment选项显示在SQL语句文本的末尾，而不管它们是在SQL命令中指定的位置。
一些显示的/*#OPTIONS */ comment选项没有在SQL命令中指定，而是由编译器的预处理器生成的。
例如 /*#OPTIONS {"DynamicSQLTypeList": ...} */

/*#OPTIONS */ comment选项显示在Show Plan语句文本、缓存的查询查询文本和SQL语句语句文本中。

为仅在/*#OPTIONS */ comment选项中不同的查询创建一个单独的缓存查询。

并行查询处理

并行查询提示指示系统在多处理器系统上运行时执行并行查询处理。
这可以极大地提高某些类型查询的性能。
SQL优化器确定一个特定的查询是否可以从并行处理中受益，并在适当的时候执行并行处理。
指定并行查询提示并不强制对每个查询进行并行处理，只强制那些可能从并行处理中受益的查询。
如果系统不是多处理器系统，则此选项无效。
要确定当前系统上的处理器数量，请使用 %SYSTEM.Util.NumberOfCPUs()方法。

可以通过两种方式指定并行查询处理:
- 在系统范围内，通过设置auto parallel选项。
- 在每个查询的FROM子句中指定%PARALLEL关键字。

并行查询处理应用于SELECT查询。
它不应用于插入、更新或删除操作。

系统范围的并行查询处理

可以使用以下选项之一来配置系统范围的自动并行查询处理:
- 在管理门户中选择System Administration，然后选择Configuration，然后选择SQL和对象设置，最后选择SQL。
查看或更改在单个进程中执行查询复选框。
注意，该复选框的默认值是未选中的，这意味着并行处理在默认情况下是激活的。
- 调用$SYSTEM.SQL.Util.SetOption()方法，如下: SET status=$SYSTEM.SQL.Util.SetOption("AutoParallel",1,.oldval).
默认值是1(自动并行处理激活)。
要确定当前的设置，调用$SYSTEM.SQL.CurrentSettings()，它会显示为%PARALLEL选项启用自动提示。

注意，更改此配置设置将清除所有名称空间中的所有缓存查询。

当激活时，自动并行查询提示指示SQL优化器对任何可能受益于这种处理的查询应用并行处理。
在IRIS 2019.1及其后续版本中，自动并行处理是默认激活的。
从IRIS 2018.1升级到IRIS 2019.1的用户需要明确激活自动并行处理。

SQL优化器用于决定是否对查询执行并行处理的一个选项是自动并行阈值。
如果激活了系统范围的自动并行处理(默认)，可以使用$SYSTEM.SQL.Util.SetOption()方法将自动并行处理的优化阈值设置为整数值，如下所示: SET status=$SYSTEM.SQL.Util.SetOption("AutoParallelThreshold",n,.oldval)。
n阈值越高，将此特性应用于查询的可能性就越低。
此阈值用于复杂的优化计算，但可以将此值视为必须驻留在已访问映射中的元组的最小数量。
默认值为3200。
最小值为0。
要确定当前的设置，调用$SYSTEM.SQL.CurrentSettings()，它显示%PARALLEL选项的自动提示阈值。

当自动并行处理被激活时，在分片环境中执行的查询将始终使用并行处理执行，而不管并行阈值是多少。

针对特定查询的并行查询处理

可选的%PARALLEL关键字在查询的FROM子句中指定。
它建议跨系统的IRIS使用多个处理器(如果适用的话)并行处理查询。
这可以显著提高使用一个或多个COUNT、SUM、AVG、MAX或MIN聚合函数和/或groupby子句的查询的性能，以及许多其他类型的查询。
这些通常是处理大量数据并返回小结果集的查询。
例如，SELECT AVG(SaleAmt) FROM %PARALLEL User.AllSales GROUP BY Region都可使用并行处理。

仅指定聚合函数、表达式和子查询的“一行”查询执行并行处理，无论是否带有GROUP BY子句。
但是，同时指定单个字段和一个或多个聚合函数的“多行”查询不会执行并行处理，除非它包含GROUP BY子句。
例如，SELECT Name,AVG(Age) FROM %PARALLEL Sample.Person不执行并行处理，但是 SELECT Name,AVG(Age) FROM %PARALLEL Sample.Person GROUP BY HomeState 执行并行处理。

如果在运行时模式下编译指定%PARALLEL的查询，则所有常量都被解释为ODBC格式。

指定%PARALLEL可能会降低某些查询的性能。
在一个有多个并发用户的系统上运行%PARALLEL查询可能会降低整体性能。

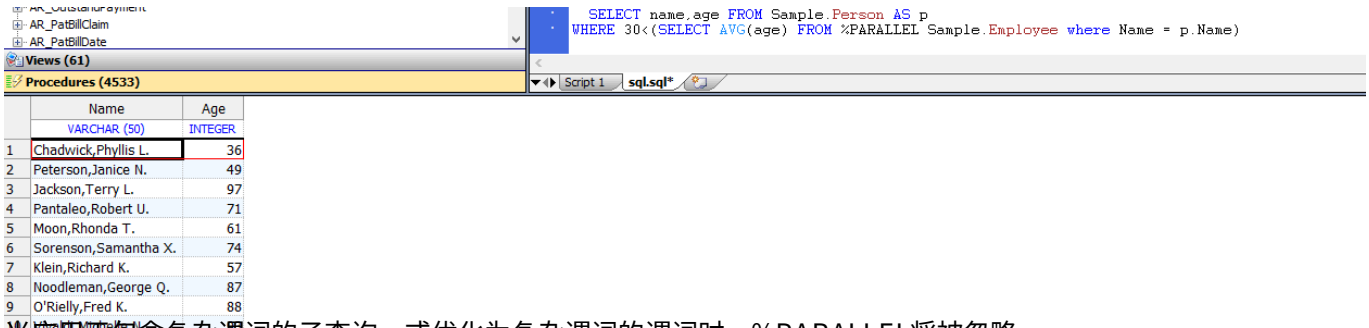
在查询视图时可以执行并行处理。
但是，即使显式地指定了%parallel关键字，也不会对指定%VID的查询执行并行处理。

%PARALLEL的子查询

%PARALLEL用于SELECT查询及其子查询。
插入命令子查询不能使用%PARALLEL。

当应用于与外围查询相关的子查询时，%PARALLEL将被忽略。
例如:

```
SELECT name,age FROM Sample.Person AS p
WHERE 30<(SELECT AVG(age) FROM %PARALLEL Sample.Employee where Name = p.Name)
```



当应用于包含复杂谓词的子查询，或优化为复杂谓词的谓词时，%PARALLEL将被忽略。
被认为复杂的谓词包括FOR SOME和FOR SOME %ELEMENT谓词。

并行查询处理被忽略

无论AUTO PARALLEL选项设置如何，或者FROM子句中是否存在%PARALLEL关键字，某些查询都可能使用线性处理，而不是并行处理。InterSystems IRIS在优化查询后决定是否对该查询使用并行处理，并应用其他查询优化选项(如果指定)。RIS可以确定优化形式的查询不适合并行处理，即使用户指定的形式的查询似乎受益于并行处理。可以使用Show Plan确定InterSystems IRIS是否以及如何对查询进行分区以进行并行处理。

在以下情况下，指定%PARALLEL不会执行并行处理。查询成功执行，没有发出错误，但没有执行并行化：
- 该查询包含FOR某些谓词。
- 该查询包含一个TOP子句和一个ORDER BY子句。
这种子句组合优化了不使用并行处理的最快时间到第一行。
添加FROM子句%NOTOPOPT optimization -option关键字可优化以最快速度检索完整结果集。

如果查询不包含聚合函数，%PARALLEL和%NOTOPOPT的组合将执行查询的并行处理。
- 包含左外连接或内连接(其中ON子句不是相等条件)的查询。例如, FROM %PARALLEL Sample.Person p LEFT OUTER JOIN Sample.Employee e ON p.dob > e.dob.
这是因为SQL优化将这种类型的连接转换为完整的外部连接。
对于完整的外部连接，%PARALLEL将被忽略。
- %PARALLEL和%INORDER优化不能同时使用;
如果两者都指定，%PARALLEL将被忽略。
- 查询引用一个视图并返回一个视图ID (%VID)。
- 如果表有BITMAPEXTENT索引，COUNT(*)不使用并行处理。
- %PARALLEL用于使用标准数据存储定义的表。
可能不支持将其与自定义存储格式一起使用。
%PARALLEL不支持全局临时表或具有扩展全局引用存储的表。
- %PARALLEL用于可以访问一个表的所有行的查询，使用行级安全(ROWLEVELSECURITY)定义的表不能执行并行处理。
- %PARALLEL用于存储在本地数据库中的数据。
它不支持映射到远程数据库的全局节点。

共享内存的考虑

对于并行处理，IRIS支持多个进程间队列(IPQ)。
每个IPQ处理单个并行查询。
它允许并行工作单元子流程将数据行发送回主流程，这样主流程就不必等待工作单元完成。
这使得并行查询能够尽可能快地返回第一行数据，而不必等待整个查询完成。
它还改进了聚合函数的性能。

并行查询执行使用来自通用内存堆(gmheap)的共享内存。
如果使用并行SQL查询执行，用户可能需要增加gmheap大小。
一般来说，每个IPQ的内存需求是4 x 64k = 256k。
InterSystems IRIS将一个并行SQL查询拆分为可用的CPU核数。
因此，用户需要分配这么多额外的gmheap:

```
<Number of concurrent parallel SQL requests> x <Number cores> x 256 = <required size
increase (in kilobytes) of gmheap>
```

注意，这个公式不是100%准确的，因为一个并行查询可以产生同样是并行的子查询。
因此，明智的做法是分配比这个公式指定的更多的额外gmheap。

分配足够的gmheap失败将导致错误报告给messages.log。
SQL查询可能会失败。
其他子系统尝试分配gmheap时也可能出现其他错误。

要查看一个实例的gmheap使用情况，特别是IPQ使用情况，请在管理门户的主页上选择System Operation，然后选择System usage，然后单击Shared Memory Heap usage链接;

要更改通用内存堆或gmheap(有时称为共享内存堆或SMH)的大小，请从管理门户的主页选择 “ 系统管理 ”，然后是 “ 配置 ”，然后是 “ 附加设置 ”，最后是 “ 高级内存 ”;

缓存查询注意事项

如果你正在运行一个缓存的SQL查询，使用%PARALLEL，当这个查询被初始化时，你做了一些事情来清除缓存的查询，那么这个查询可能会从一个工人作业报告一个<NOROUTINE>错误。
导致缓存查询被清除的典型情况是调用$SYSTEM.SQL.Purge()或重新编译该查询引用的类。
重新编译类将自动清除与该类相关的任何缓存查询。

如果发生此错误，再次运行查询可能会成功执行。

从查询中删除%PARALLEL可以避免出现此错误。

SQL语句和计划状态

使用%PARALLEL的SQL查询可以产生多条SQL语句。
这些SQL语句的计划状态是Unfrozen/Parallel。
计划状态为 “ 已冻结 ” / “ 并行 ” 的查询不能通过用户操作进行冻结。

生成报告

可以使用生成报告工具向InterSystems Worldwide Response Center (WRC) customer support提交查询性能报告，以便进行分析。
可以使用以下任意一种方式从管理门户运行生成报告工具:

1. 必须首先从WRC获得WRC跟踪号。可以使用每个管理门户页面顶部的Contact按钮从管理门户联系WRC。在WRC编号区域中输入此跟踪编号。可以使用此跟踪编号来报告单个查询或多个查询的性能。
2. 在 “ SQL语句 ” 区域中，输入查询文本。右上角将显示一个X图标。可以使用此图标清除SQL语句区。查询完成后，选择保存查询按钮。系统生成查询计划并收集指定查询的运行时统计信息。无论系统范围的运行时统计信息设置如何，生成报告工具始终使用收集选项3：记录查询的所有模块级别的统计信息进行收集。由于在此级别收集统计信息可能需要时间，因此强烈建议您选中 “ 在后台运行保存查询进程 ” 复选框。默认情况下，此复选框处于选中状态。

当后台任务启动时，该工具显示 “ 请等待…… ” ，禁用页面上的所有字段，并显示一个新的视图进程按钮。
单击View Process按钮将在新选项卡中打开Process Details页面。
在流程详细信息页面，您可以查看该流程，并可以 “ 暂停 ” 、 “ 恢复 ” 或 “ 终止 ” 该流程。
进程的状态反映在Save查询页面上。
当流程完成时，当前保存的查询表将被刷新，View process按钮将消失，页面上的所有字段将被启用。

3. 对每个查询执行步骤2。
   每个查询将被添加到当前保存的Queries表中。
   注意，该表可以包含具有相同WRC跟踪号的查询，也可以包含具有不同跟踪号的查询。
   完成所有查询后，继续步骤4。

对于列出的每个查询，可以选择Details链接。
该链接将打开一个单独的页面，其中显示完整的SQL语句、属性(包括WRC跟踪号和IRIS软件版本)，以及包含每个模块的性能统计信息的查询计划。

- 要删除单个查询，请从 “ 当前保存的查询 ” 表中选中这些查询的复选框，然后单击 “ 清除 ” 按钮。
- 要删除与WRC跟踪编号关联的所有查询，请从当前保存的查询表中选择一行。WRC编号显示在页面顶部的WRC编号区域。如果您随后单击清除按钮，则对该WRC编号的所有查询都将被删除。

4. 使用查询复选框选择要报告给WRC的查询。要选择与WRC跟踪编号关联的所有查询，请从当前保存的查询表中选择一行，而不是使用复选框。在这两种情况下，都可以选择Generate Report按钮。生成报告工具创建一个XML文件，其中包括查询语句、具有运行时统计信息的查询计划、类定义以及与每个所选查询相关联的SQL int文件。

如果选择与单个WRC跟踪编号关联的查询，则生成的文件将具有默认名称，如WRC12345.xml。如果选择与多个WRC跟踪编号关联的查询，则生成的文件将具有默认名称WRCMultiple.xml。

将出现一个对话框，要求指定保存报告的位置。保存报告后，可以单击Mail to链接将报告发送给WRC客户支持。使用邮件客户端的附加/插入功能附加文件。

#SQL #Caché #InterSystems IRIS #InterSystems IRIS for Health

源URL:
https://cn.community.intersystems.com/post/%E7%AC%AC%E4%BA%94%E7%AB%A0-%E4%BC%98%E5%8C%96%E6%9F%A5%E8%AF%A2%E6%80%A7%E8%83%BD%EF%BC%88%E5%9B%9B%EF%BC%89