SYBASE
技术服务园地
连载（74）

解决方案

Sybase 15.3 横向扩展查询性能

使用PlexQ分布式查询平台、全共享的MPP架构（三）

（上接第9期）

4.1 DQP的可扩展性

一个查询可能只有全面并行化并充分利用单个节点的CPU资源，才能从DQP获益。另外，

Sybase IQ主存储和共享临时存储必须不能受到I/O的限制。

DQP使用逻辑服务器上所有节点的可用的内存和CPU资源。一般来讲，可用的节点和资源越多，查询性能越高。基于任务单元的数量，存在一个上限。如果没有足够的任务单元传送到Multiplex中所有可用的CPU上，则只有一部分CPU被使用。逻辑服务器中节点的当前工作负载显然将影响性能。

分配更多的内存到临时缓存将使基于哈希的算法更可能扩展。一个大的临时缓存比一个大的主缓冲对DQP更重要。共享临时存储的I/O带宽—用于分配任务和传递即时结果—对分布式查询的性能非常关键，因此如果你的存储层提供了分层性能特性，将IQ_SHARED_TEMP置于快速的存储将产生更好的结果。

这看起来似乎显而易见，但是所有的分布式碎片必须在最终的结果集被生成以及返回给请求的应用之前全部处理完毕。因此，应该注意“最慢的碎片执行”将限制查询的整体性能。另外，尽管查询在Sybase IQ 15.3的DQP层之内自动进行分布和负载均衡，但是将负载均衡跨Multiplex连接以将更多密集型的Leader节点任务分散到Multiplex中的所有节点上，这仍是一个好主意。

4.2 最有可能从DQP获益的查询

DQP用于报表高度密集的Multiplex环境。加载性能不受DQP选项的影响，尽管加载可通过配置多个Multiplex写节点而并行化。同时，当内存和CPU在Multiplex间被平衡时，DQP将获得更好的执行。

某些类型的查询将比其他查询更具扩展性。更可能被很好分布的查询具有以下属性：

（1）计算密集型的列扫描，比如LIKE条件。

（2）包含汇总、繁多的表达、数值型数据类型的复杂查询。

（3）由可减小中间和最终结果的查询碎片组成的查询。一个例子就是，一系列在顶部带有“group by hash”的hash joins。

（4）经常使用基于哈希处理的低基数数据，更可能扩展。这常常出现在星型模型中，其特征是一个大的事实表和一些拥有低基数维度的表。

（5）如果你有中等基数的表，你可以对数据库选项进行调优，分配更多的内存到临时缓冲，偏置查询优化器去选择更多基于哈希的算法。

4.3 通常有可能从DQP获益的查询

正如前面所述，有些类型的查询天然的不能被很好扩展，查询优化器也决定不对这些查询进行分布，因为它们在单个节点上可能执行得更好。这些类型的查询具有的特征包括：

（1）返回大量行的查询，因此返回行占用了查询执行时间的大部分。注意，从一个查询的“顶部”生成行是一个系列操作，不能被分布。

（2）小查询：用时不超过2 s的查询不可能从DQP中获益。2 s～10 s的查询也不太可能获益。超过10秒的查询更可能从中获益。

（3）有大量碎片的查询。如果有大量碎片，这通常意味着包含排序。这可能导致无法扩展，因为对大量数据排序，要使用IQ_SHARED_TEMP DBSpace中的磁盘存储执行这个排序。这也是为什么共享临时DBSpace应该尽可能被放置到快速的存储上的原因。

（4）join高基数、大表，将导致合并joins。这不能像hash joins一样扩展。

4.4 你可以做什么影响DQP的可扩展性

有多种服务器和数据操作可影响查询的并行化和性能。

（1）Max_Query_Parallelism：这个数据库选项设定了一个上限，限制了优化器将如何允许查询操作并行，比如Joins，Group By以及Order By。缺省的值是64。有超过64颗CPU核的系统通常可以从更大的值中获益一直到系统中的全部CPU数，最大值为512。

（2）Minimize_Storage：在向表中加载数据前设置该数据库选项为“on”，或者在列定义上使用

IQ_UNIQUE。使用了参照表的FP（1），FP（2），FP（3）索引将代替flat FP索引而被创建。这占用更少的空间并减少I/O（尽管FP(3)索引消耗大量内存，因此应审慎的使用它们）。

（3）Force_No_Scroll_Cursors：如果你不需要向后回滚游标，将该数据库选项设置为“on”以减少临时存储需求。

（4）Max_IQ_Threads_Per_Connection：控制每个连接的线程数。对大的系统，你会发现通过提高该值所带来的性能优势。

（5）Max_IQ_Threads_Per_Team：控制分配给单个操作（比如一个列上的LIKE谓词）执行的线程数。对于大的系统，你会发现通过提高该值所带来的性能优势。

（6）Max_Hash_Rows：将该数据库选项设置为主机上的每4 GB RAM 250万。例如，在一个64 GB的系统上设置为4 000万。这会鼓励查询优化器使用更好扩展的基于哈希的join和group by算法。然而，在此有个警告：对于超大的哈希表，分布可能会带来性能的倒退，由于将哈希表从一个节点取出并在另一个节点重组它们所需的时间。DQP将试图弥补这种情况，当哈希表变得非常大时，即使内存可以满足，也不去分布基于哈希的操作。

（7）-iqgovern：这个服务器选项设定了一个特定服务器上并发查询的量。通过设定-iqgovern开关，你可以帮助IQ维持吞吐量，给查询充足的资源快速完成。缺省的值是2倍的CPU数+10。对于有大量活动连接的点，你可能需要将这个值设得更低。

（8）-iqtc：这个数据库选项设置临时缓冲大小。临时缓冲既被本地临时存储也被共享临时存储使用。DQP必须利用IQ_SHARED_TEMP来执行处理，因此要求充足的临时缓冲。你可能需要分配比DQP负载主缓冲更多的内存给它。

（9）同时，有两个特别为DQP提供的数据库选项：

a.MPX_Work_Unit_Timeout：当一个Worker节点无法在mpx_work_unit_timeout时间内完成查询碎片的处理，该任务将返回到Leader节点重试。如果你发现timeout出现并对DQP 的性能产生负面影响，你可以增大timeout的值允许Worker完成任务。尽管一般而言，你不可能遇到timeout问题，除非你有一些其他的底层问题。

b.DQP_Enabled：这是一个让你为数据库连接设置的选项。如果DQP出现，但是你没有看到它带来的好处，你可以关掉它。

4.5 设置共享临时存储的大小

一个位于高速存储硬件上的充足的共享临时空间对分布式查询的性能至关重要。尽管很难提前计算分布式查询需要多大的共享临时存储，但是也有一些已经被发现的趋势：

（1）当一个分布式查询被执行的时候，共享临时空间的使用在Multiplex中的节点间变化很大。

（2）共享临时空间的使用并不与查询的可扩展性相关。那些不能很好扩展的查询相比于可以很好扩展的查询，可能会使用相同或更多的共享临时空间。

（3）那些在单个节点上使用更多临时缓冲/空间的查询，当运行分布式时一般也会使用更多的共享临时空间，但是并没有明显的倍数关系。

（4）跨Multiplex使用的共享临时空间的最大值固定不变，不管执行一个特定的分布式查询的节点数是多少。

（5）一个节点上所要求的共享临时空间的大小随着执行同一个分布式查询的并发用户数而增加。换句话说，更高的工作负载要求更多的共享临时存储。

确保你有可用的存储添加到共享临时存储，如果你发现它的分配不是很适合。你可以动态的添加空间，无需停止IQ 服务器。

4.6 DQP 单一的查询负载测试结果

一个分布式查询的性能变化显著的依赖于这个查询本身、以及执行它的Sybase IQ Multiplex的配置和工作负载。下面的结果是到目前为止在可控制的、内部测试环境中所取得的结果。

这些测试（由单一客户端发起的单一的大的查询）运行于Sybase IQ Multiplex之上，配置如下：

（1）Dell Blade M1000E,Power Edge Enclosure-16 X M610 Blade服务器；56XX处理器（224-8593）

（2）2 x quad-core（Intel Xeon E5620 2.4 Ghz）

（3）48 GB内存

（4）2 x 300 GB SAS Drives (Raid)

（5）Dual-Channel 8 Gbps Fibre HBA

（6）Dual-Port 10GbE Network Card

（7）2 x Fiber Switch-Brocade M5424 FC8 Switch+AG，24 ports

（8）10 GB Private Network NFS Server-Del R710

（9）quad-core

(10) 24 GB 内存

(11)8x1TB Near-Line SAS Drives

(12) 存储

- 6 x PAC Storage 12-Bay 4 GB Dual Raid Control ers w/12 x 300GB 15K SAS Drives

- 6 x PAC Storage 12-Bay EBOD(Expansion Shelves) w/12 x 300 GB 15 K SAS Drives

- RAID-o striping with LUN stripe size = 64 KB

下面的每个测试都显示了 Leader 节点对特点查询的查询计划，一个柱状图显示了从 1 个到 8 个服务器的性能扩展。查询的名字没有特别的意义，仅仅是唯一的标识它。在查询计划中，注意“3 条竖线”的注解说明了查询处理的分布，如图 8、图 9、图 10。

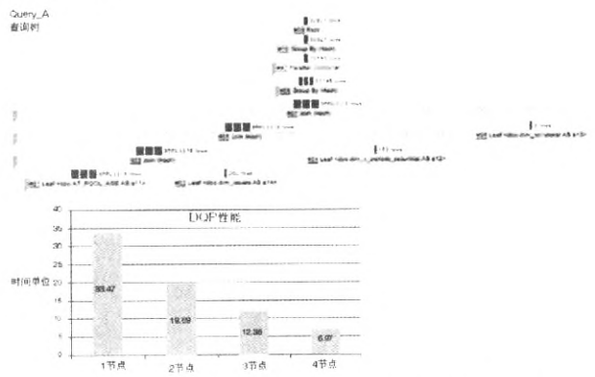

图 8　Query_A 从 1 个到 8 个 Multiplex 节点扩展

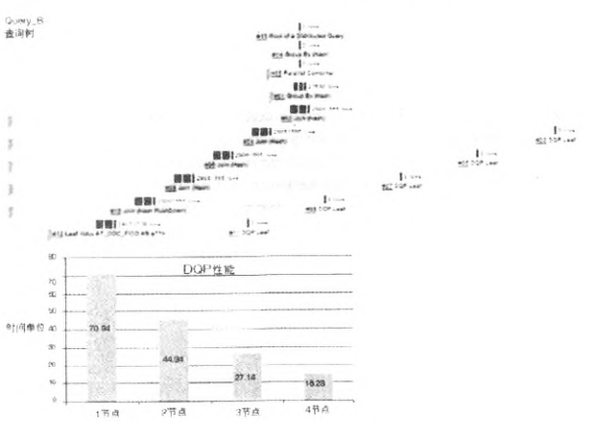

图 9　—Query_B 从 1 个到 8 个 Multiplex 节点扩展

5　结束语

本文已经给了你一个关于 PlexQ — Sybase IQ 15.3 中新引入的一组令人激动的功能概览，包含了旨在提供一个高性能、资源效率、以及简化操作的平台

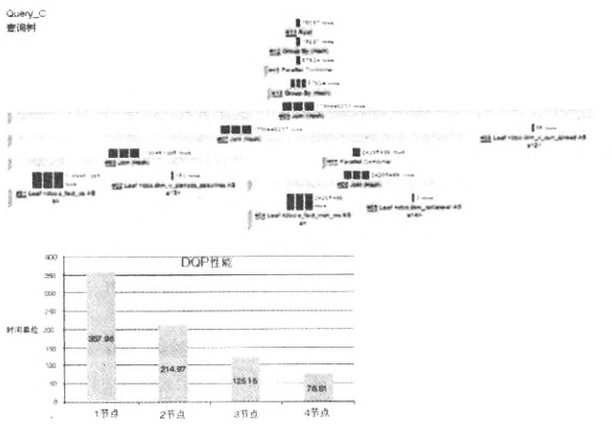

图 10　—Query_C 从 1 个到 8 个 Multiplex 节点扩展

的分布式查询处理。DQP 的设计是为了充分利用 Sybase IQ Multiplex 的 CPU 资源以扩展大的、复杂的、与 CPU 绑定的查询的性能。DQP 可以通过分解查询并将查询碎片在多个 Sybase IQ 服务器上进行分布以并行化执行，从而大幅提升查询性能。这个新的功能推动了 Sybase IQ 平台迈向一个可以进一步利用分布式资源获得更好的查询性能和资源效率的“全共享的 MPP”架构。在今天不断变化的、复杂的、高度竞争的世界里，快速回答时间关键型问题是企业成功的法宝。

文 / 赛贝斯软件（中国）有限公司

（完）