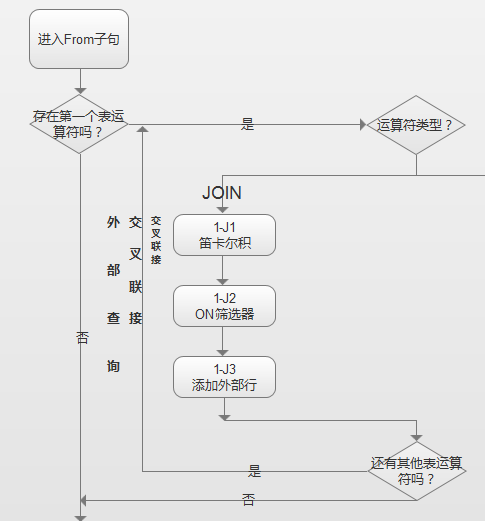
**逻辑查询处理**

SQL与其他编程语言不同的最明显特征是代码的查询顺序。在大多数语言中，代码是按照编写顺序来处理的；但在SQL中，第一个要处理的子句是FROM子句，尽管SELECT是第一个出现，但它几乎都是最后才处理。

每一步都会生成一个虚拟表，该虚拟表会称为下一步的输入，这些虚拟表对于用户是不可见的。如果在查询中没有指定某一个子句，则会跳过相应的步骤。

**步骤1：From阶段**

****

From阶段负责标识要查询的表，如果指定了表运算符，这个阶段还要按照从左到右的顺序对这些运算符进行处理。

**步骤1-J1：执行笛卡尔积（交叉联接）**

这个阶段对联接涉及的两个表执行笛卡尔积运算，生成虚拟表**VT1-J1**。这个虚拟表为左表行和有表行的每一个可能的组合包含一行数据。如果左表包含n行，右表包含m行，**VT1-J1**包含m \* n行。

**步骤1-J2：应用ON筛选器（联接条件）**

ON筛选器是查询中可以指定的三种筛选器（ON，Where，Having）中的第一个。ON筛选器作用于虚拟表（VT1-J1）中的所有行。只有使<on\_predicate>为True的那些行，才会包含在这一步返回的虚拟表（**VT1-J2**）。

**步骤1-J3：添加外部行（Outer Row）**

这一步只在外联接（outer join）中发生。对于外联接，通过为其指定一种外联接类型（LEFT ,RIGHT或FULL），就把一个或两个表标记为保留表（preserved table）。把表标记为**保留表**，即表示希望返回该表的所有行，即使ON筛选器过滤掉了这些行。

**左外联接**是把左表标记为保留表，**右外联接**是把右表标记为保留表，**完全外联接**是把两个表标记为保留表。

步骤1-J3返回VT1-J2中的行，以及保留表在1-J2中过滤掉的行。这些新行被称为**外部行**（outer row）。外部行中非保留表的列值被赋值为NULL，最后生成**VT1-J3**。

注意：*如果From子句中有多个表运算符，则按照从左到右的顺序进行处理。每个表运算符的结果作为下一个表运算符的左输入。最终生成的虚拟表将作为下一阶段的输入。*

**步骤2：Where阶段**

对上一步返回的虚拟表中返回的所有行应用Where筛选器。只有让<where\_predicate>逻辑条件为True的行，才会组成这一步要返回的虚拟表**VT2**。

注意：由于还有对表进行分组，所以在Where子句中不能使用聚合。

对于包含外联接的查询，有一个问题是让人困惑的，**何时使用ON筛选器？何时Where筛选器？**

两者的主要区别是：ON在添加外部行之前执行，而Where在添加外部行之后执行。ON中筛选掉的行不是最终的，因为步骤1-J3会将这些行再添加回来；相反Where筛选器对行的删除是最终的。

需要注意的是，只有在使用外联接时，ON和Where才有这种逻辑区别。当使用内连接时，它们是一致的。

**步骤3：GROUP BY阶段**

在Group By阶段，根据<group\_by\_specification>指定的列表，将上一步返回的虚拟表中的行分配到各个组，这一组列被称之为**分组集**。

*例如：GROUP BY C.customerid中C.customerid就是一个分组集。*

这个阶段将上一步返回的虚拟表中的行按组进行分组。由分组集中的所有属性值的每个唯一组合标识出一个组。最终得到的虚拟表VT3由两部分组成：分组后原来的行(原始信息)，组标识符（组信息）。

|  |  |  |  |
| --- | --- | --- | --- |
| 组信息  C.customerid | 原始信息  C.customerid | C.city | O.orderid |
| Jack | Jack  Jack | ShangHai  ShangHai | 1  2 |
| Lucy | Lucy  Lucy  Lucy | Chongqing  Chongqing  Null | 3  4  5 |
| Ajax | Ajax | Wuhan | 6 |

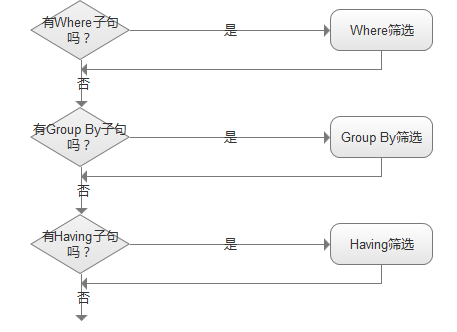
最终，Group By子句的查询将为每一个组生成一行。所以在查询中指定了Group By子句，则后面的所有步骤（Having，Select等）只能指定每个分组上的标量（单个）值表达式。这些表达式可以是Group By列表中的列，表达式（C.customerid），以及聚合行数（Count(C.customerid)）。

这个阶段会认为两个NULL值是相等的。也就是说，所有的NULL值将被分配到同一个组中。

**步骤4：HAVING阶段**

HAVING筛选器用于对上一步返回的虚拟表中的组进行筛选。只有使<having\_predicate>逻辑条件取值为True的组，才会称为这一步返回的虚拟表（VT4）的一部分。

注意：子查询不能作为聚合函数的输入。



**步骤5：SELECT阶段**

虽然SELECT子句出现在查询的最前面，但却放在第五步处理。

**步骤5-1：计算表达式**

SELECT列表中的表达式可以返回上一步得到的虚拟表的基础列，也可以是对这些基础列的操作。如果查询是一个聚合查询，就只能引用上一步返回的虚拟表中的组信息。如果要引用原始信息，就必须对它们进行聚合运算。这个阶段得到的是虚拟表**VT5**。

**步骤5-2：应用Distinct子句**

如果在查询中指定了Distinct子句，将从上一句返回的虚拟表中删除重复的行，并生成虚拟表**VT5-2**。

**步骤5-3：应用Top选项**

Top选项是T-SQL特有的一项功能，允许指定要返回的指定行数或百分比，根据查询的Order by子句来选择指定的数量的行。最终生成虚拟表**VT5-3**。

如果没有指定Order By子句或者未指定With Ties选项，或是根本没有指定Order By子句，那么Top查询就是非确定的。也就是说，返回的行只是SQL Server碰巧在物理上最先访问到的行，因此可能会产生不同的结果，但都可以认为是正确的。

**步骤5：ORDER BY阶段**

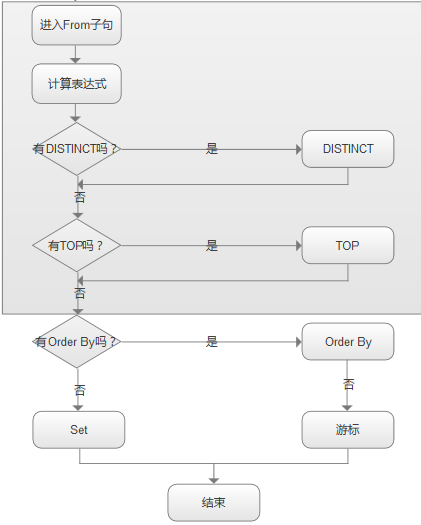
这一步按Order By子句中的列名列表对上一步返回的行进行排序，返回游标**VT6**。Order By子句是唯一可以重用SELECT列表中创建列名的步骤。

如果指定了Distinct，则Order By子句中的表达式只能访问上一步返回的虚拟表VT5。如果没有指定Distinct则可以指定任何在Select子句中的表达式。

如果查询中带有Order By子句，则不能用这样的查询来定义表表达式。表表达式包括：视图，内联表值函数，派生表和公用表表达式（CTE）。

除非真的需要有序行，否则不要指定Order By子句，因为排序需要耗费一定的资源。SQL Server需指定有序的索引扫描或sort运算符。

Order By子句认为两个NULL值是相等的。也就是说所有的NULL值是会排序到一起。T-SQL认为NULL值比已知值小。



查询优化

在处理性能问题时，数据库专家倾向于关注系统的技术层面，如资源队列、资源利用率等。而用户只把性能问题简单地认为是等待时间，他们发出一个请求，然后等待返回结果。用户通常认为 在交互请求后，超过三秒才得到响应，就算存在性能问题了。他们并不真关心平均每个磁盘转轴上 有多少个命令在等待，或者缓存命中率（cache hit ratio)是多少，也不关心阻塞，CPU利用率，缓 存中数据页的平均停留时间（page life expectancy)等。他们只关心等待时间，这正是优化性能应该着手的地方.

我推荐的一种使用自顶向下方法的优化论。这种方法首先分析实例级的等待时间，再通过一系 列步骤将其不断细化，直到找出系统中导致大量等待的进程组件。一旦找出这些令人讨厌的进程, 就可以集中优化它们了。以下是这种方法论的主要步骤：

1. 分析实例级别等待
2. 关联等待和队列
3. 确定方案
4. 细化到数据库/文件级
5. 细化到进程级
6. 优化索引/查询

实例级别等待

优化方法论的第一步是在实力级别上找出什么类型的等待占用了大部分的等待时间。这可以通过查询动态管理图(DMV,dynamic management view) sys.dm\_os\_wait\_stats来完成。该DMV包含了400多种等待类型，大多数类型至少在SQL Server联机丛书中至少有一段文字说明。可以将这个DMV作为优化的起点。而其他一些优化工具则不适合在开始使用。因为它们提供的信息量太大，让你无法入手，很容易迷失其中。

SELECT

wait\_type, --等待类型

waiting\_tasks\_count, --该类型等待的数量

wait\_time\_ms, --该类型总的等待时间

max\_wait\_time\_ms, --最大等待时间

signal\_wait\_time\_ms --从收到信号到开始运行之间的时间差

FROM sys.dm\_os\_wait\_stats

ORDER BY wait\_type;

这个DMV会在服务器最后一次重启时重置。

Signal wait time ms 属性就是表示从线程收到资源可用的信号开始，到线程得到CPU，开始使用资源的为止经历的时间。可以想到，这个数值很高，就说明CPU存在问题。

在各种等待类型中，可以找到与锁、闩锁（latch，一种轻型锁）、I/O (包括I/O闩锁）、并行查 询（parallelism)、事务口志、内存、编译、OLEDB (链接服务器和其他OLEDB组件）等相关的等待。

通常，你会忽略某些类型的等待，例如，睡眠等待类型 (SLEEP，当线程被挂起，未执行任何操作时发生)、队列等待类型（QUEUE，当工作线程空闲，等待分配任务时发生），以及在SQL Server联机丛书中特别描述的一 些并不代表出现问题的等待类型（如 CLR\_AUTO\_EVENT、REQUEST\_FOR\_DEADLOCK\_ SEARCH 等等）。要确保过滤掉不相关的等待.以免它们影响你的计算。

**等待的类型**

**资源等待** 当工作者请求访问由于某个其他工作人员正在使用该资源或尚未可用的资源而无法访问的资源时，会发生资源等待。资源等待的示例是锁，锁存器，网络和磁盘I / O等待。锁定和锁存等待是等待同步对象

**队列等待**  
队列等待工作空闲时等待分配工作。队列等待最常见于系统后台任务，例如死锁监视器和已删除的记录清理任务。这些任务将等待将工作请求放入工作队列。即使没有新的数据包放入队列，队列等待也可能周期性地变为活动状态。

**外部等待**  
当SQL Server工作程序等待外部事件（如扩展存储过程调用或链接服务器查询）完成时，会发生外部等待。当您诊断阻塞问题时，请记住外部等待并不总是暗示工作者是空闲的，因为工作人员可能会主动运行某些外部代码。

|  |  |  |
| --- | --- | --- |
| 类型 | 描述 | 影响 |
| LAZYWRITER\_SLEEP | 在惰性写入器任务被挂起时发生。这是等待的后台任务所花费的时间的度量。当您在寻找用户档位时，请不要考虑此状态。 | 无 |
| SQLTRACE\_INCREMENTAL\_FLUSH\_SLEEP | 此等待类型是线程在检查SQL跟踪信息以刷新到跟踪文件之间休眠的时间 | 无 |
| SLEEP\_TASK | 在等待发生通用事件时任务休眠时发生。 | 无 |
| SLEEP\_SYSTEMTASK | 在等待tempdb完成启动时在后台任务启动期间发生。 | 无 |
| FT\_IFTS\_SCHEDULER\_IDLE\_WAIT | 此等待类型是线程正在等待在全文搜索后台任务队列中输入某些工作的时间。 | 无 |
| SERVER\_IDLE\_CHECK | 在资源监视器尝试将SQL Server实例声明为空闲或尝试唤醒时，在SQL Server实例空闲状态同步期间发生。 | 无 |
| CLR\_AUTO\_EVENT | 当某任务当前正在执行公共语言运行时 (CLR) 执行并且正在等待特殊的自动事件启动时出现。 通常会出现长时间等待，这并不意味着出现问题。 | 无 |
| REQUEST\_FOR\_DEADLOCK\_SEARCH | 在死锁监视器等待开始下一次死锁搜索时出现。 在两次死锁检测之间可能出现该等待，长时间等待此资源并不指示出现问题。 | 无 |
| SQLTRACE\_BUFFER\_FLUSH | 当某任务正在等待后台任务将跟踪缓冲区每隔四秒刷新到磁盘时出现。  适用于: SQL Server 2008 R2仅。 | 无 |
| IOLATCH相关 | 在任务等待 I/O 请求中缓冲区的闩锁时发生。  长时间的等待可能指示磁盘子系统出现问题。 | 有 |
| ASYNC\_NETWORK\_IO | 当任务被阻止在网络之后时出现在网络写入中。 验证客户端是否正在处理来自服务器的数据。 | 有 |
| BROKER\_TO\_FLUSH | 此等待类型是指线程正在等待写入非活动传输对象（跟踪对话框的消息传输）到tempdb工作表以节省内存。  这种等待类型是我通常在进行等待统计分析时过滤掉的良性等待类型。此等待类型的等待时间将每1秒增加1秒。 | 无 |
| PAGEIOLATCH\_SH | 在任务等待 I/O 请求中缓冲区的闩锁时发生。 闩锁请求处于“共享”模式。 长时间的等待可能指示磁盘子系统出现问题。 | 有 |
| CMEMTHREAD | 当某任务正在等待线程安全内存对象时出现。 当多项任务尝试分配来自同一个内存对象的内存而导致出现争用时，便可能延长等待时间。 | 有 |
| CXPACKET | 当同步查询处理器交换迭代器，并生成和使用行时出现并行查询计划。 如果等待太久，无法通过优化查询（如添加索引）来减少等待时间，请考虑调整并行度的开销阈值或降低并行度。 | 有 |
| WRITELOG | 此等待类型是指线程正在等待异步I / O将日志块写入磁盘。  在以下情况下将日志块写入磁盘：   * 事务提交（除非在SQL Server 2014及更高版本中设置为延迟持久） * 日志块的最大大小为60Kb * 正在将数据文件页面写入磁盘，并且预写日志记录强制将当前日志块强制到磁盘（因为日志块包含描述正在写出的数据文件页面的更改的最新日志记录） * 该*sp\_flush\_log*  PROC执行（在SQL Server 2014以上，对于延迟耐久性日志管理）   将*WRITELOG*  视为服务器上最重要的等待之一是非常普遍的  ，但是你仍然可以做很多事情来尝试缓解它们。  您可以做很多事情来减少  *WRITELOG*等待和等待时间，包括：   * 减少生成的事务日志量 * 减少日志刷新的频率 * 减少/消除同步HA技术无意中听到的日志刷新 * 对于可用性组，远程日志副本会导致[*HADR\_SYNC\_COMMIT*](http://www.sqlskills.com/help/waits/hadr_sync_commit/)等待 * 升级到2012或更高版本，将max-outstanding-log-writes从32提高到112 * 将事务日志放在I / O子系统的最快部分 * 考虑实施延迟耐久性（在2014+）或内存中的OLTP / Hekaton   参考：<https://www.sqlskills.com/help/waits/writelog/> | 有 |
| OLEDB | 这种等待类型是一个线程正在等待来自OLE DB提供程序的数据，该数据库在内部用于DBCC CHECK \*命令和DMV之类的东西，外部用于链接服务器通信和某些SSIS包之类的东西。 | 有 |
| LCK\_M\_IS | 此等待类型是指线程正在等待获取资源上的Intent Shared锁，并且在资源上向不同线程授予不兼容模式的至少一个其他锁。  有关排除锁定等待的一般指导：   * 无法从*sys.dm\_os\_wait\_stats*输出中确定锁资源  。您可以从  *sys.dm\_os\_waiting\_tasks*（使用  [我的脚本](http://www.sqlskills.com/blogs/paul/updated-sys-dm_os_waiting_tasks-script-2/)）或查看*sys.dm\_tran\_locks*的  *resource\_description*字段中*查看资源*，其中  *request\_status*是'WAIT'。 * 您可以使用阻止的进程报告来获取有关等待指定阈值锁定的查询的更多详细信息（请参阅[此处](http://michaeljswart.com/2011/05/when-to-use-blocked-processes-reports/" \t "_blank)）。 * 看看阻塞链的头部（即持有锁定所有人的锁的线程）使用脚本（大量在线可用 - 我没有首选的）。那个帖子在等什么？修复等待可能有助于解决阻塞问题。例如，一个线程可能持有锁并提交事务，但是有一个带有慢速I / O子系统的同步镜像，因此镜像日志写入需要很长时间，使得事务提交需要更长时间，并且锁需要更长的时间释放，导致阻塞。 * 查找锁定升级，其中*UPDATE*事务已升级为表X锁定，从而导致广泛阻塞。 * 查找导致表锁定的索引操作，并考虑使用联机索引操作（或者如果已经使用它们，请考虑*2014*年以上的*WAIT\_AT\_LOW\_PRIORITY*功能）。 * 查找指定*TABLOCK*（导致表共享锁）或*TABLOCKX*（导致表独占锁）提示的代码。 * 查找将导致获取锁定然后等待用户输入或长时间未提交事务的应用程序代码。 * 请考虑创建非聚簇索引以从基础堆/聚簇索引中删除行锁。 * 考虑使用快照隔离或读取提交的快照隔离，以允许读者不采用S / IS锁定并减少阻塞。 * 检查正在使用的正确隔离级别为*REPEATABLE\_READ*，*SERIALIZABLE*将保持S / IS锁定，直到事务结束。 * 检查是否意外使用*SERIALIZABLE*隔离级别，使用分布式事务或错误作用域的.Net *TransactionScope*对象。   *LCK\_M\_IS*的具体指导等待：   * 对于Intent Shared锁，资源可以是页面，分区或表。 * 常见的阻止程序是发生锁升级的表X（独占）锁，或索引构建/重建的SCH\_M（模式修改）锁。 | 有 |
| LCK\_M\_IX | 当某任务正在等待获取意向排他 (IX) 锁时出现。  LCK\_M\_IX的具体指导 等待：   * 对于Intent Exclusive锁，资源可以是页面，分区或表。 * 常见的阻止程序是发生锁升级的表X（独占）锁，或索引构建/重建的SCH\_M（模式修改）锁。 * 如果阻止程序持有表S锁，请调查阻塞线程为何具有该锁（例如，在限制性隔离级别中使用TABLOCK提示或锁升级）。 | 有 |

**性能瓶颈**

在很多情况下，你会发现与**I/O相关**的等待是最常见的等待（例如，IOLATCH等待〉，其中有 几个原因。I/O通常是数据处理操作所涉及的最昂贵的资源。而且，当査询或索引没有经过良好地设计或优化时，结果一般会造成大量的I/O。此外，当客户在考虑计算机的性能时，他们通常只关注CPU 和内存，而不会对I/O子系统予以足够的重视。其实，数据库系统需要非常强健的I/O子系统。

对于与**网络相关**的等待（例如，ASYNC\_NETWORK\_IO),它们的值过高，则表明可能存在 网络问题。不过，这也可能表明客户端没有足够快地使用SQL Server发送给它的数据。

—些系统不需要访问数据的大部分.只需要非常频繁地访问数据的一少部分。联机事务处理(OLTP)环境通常就是这样的情况，它们的一些存储过程只访问数据的一小部分，但调用的次数非常频繁。在这样的环境中.代码的编译和重新编译可能就是产生瓶颈的主要原因，这时你可能会发现**信号等待（与CPU有关）**的值非常高。如果使用大量的特定査询，而不是用存储过程 和预先编译好的语句，就会因为这些特定的执行计划而消耗大最内存，这时你通常会发现 CMEMTHREAD等待类型的值非常高（当某任务正在等待线程安全的内存对象时出现这种等待）。

对于使用了太多**线程的并行查询**计划，也可能会出现问题。为了等待其他线程完成它们的工作 (CXPACKET等待），一些线程可能会等待很长时间，才可以继续工作；系统作为一个整体可能 无法提供域佳的吞吐量。这时可以考虑降低系统的最大并发度。不过，也要注意，CXPACKET等待类型有时也是其他原因造成的症状（例如，因为缺乏适当的索引而引起的过度的I/O),这时你 将会发现与I/O相关的等待存在较高的值。

OLTP系统还涉及大最小范围内的数据修改.在这种环境下事务日志志经常会成为一个瓶颈。当 SQL Server不能足够快地写完日志时，你通常会看到WRITELOG等待类型中有较高的值.

因为所有的临时表都在tempdb数据库中创建（无论是由执行计划隐式创建还是显式创建）, 所以tempdb数据库也吋能成为一个作常严重的瓶颈，SQL Server还利用tempdb的空间执行一些其他操作。tempdb数据库中的性能问题可能会导致与I/O相关的等待或其他等待出现较高的值。闩锁等待（例如.PAGE\_LATCH\_UP)出现较高的值，则表示在内部结构（如IAM、GAM、SGAM和 PFS页面）上出现了争用.原因可能是为临时表频繁地分配页面、向堆空间插入人置数据等。不正确的文件布局（layout）也可以导致这样的争用。

OLEDB等待类型代表与服务器调用、BULK INSERT、Full Text等相关的等待。不过，因为不 能生成OLEDB调用；因此，当调用开始时，就开始等待，当调用结束时，就结束等待。这意味着，这种等待类型中的较高的值不一定表示存在性能问题。

偶尔，你也会发现系统存在一呰与并发相关的（阻塞）问题.在这种情况下，锁等待（LCK) 的值将会比较高。