# 背景

数据操纵是数据库管理系统中一种最基本的操作，这种操作包括查询、插入、删除和修改等，其中，查询操作称为查询处理。

查询的执行，就是查询处理的过程，即数据库按用户指定的SQL语句中的语义，执行语义所限定的操作。但SQL语句的执行效率对数据库的效率影响较大。为了提高查询语句的执行效率，对查询语句进行优化是必不可少的。

对查询语句进行优化的技术就是查询优化技术，运用查询技术实现数据操纵功能的过程是确定给定查询的高效执行计划的过程。所谓执行计划就是查询树，它由一系列内部的操作符组成，这些操作符按一定的运算关系构成查询的一个**执行方案**。

# 概述

查询优化的追求目标，就是在数据库查询优化引擎生成一个执行策略的过程中，尽量使查询的总开销（总开销通常包括IO、CPU、网络传输等）达到最小。数据库查询优化技术主要包括查询重用技术、查询重写规则、查询算法优化技术、并行查询优化技术、分布式查询优化技术及其他方面（如框架结构）的优化技术，这6项技术构成了一个“广义的数据库查询优化”的概念。

从优化的内容角度看，查询优化又分为**代数优化**和**非代数优化**，或称为**逻辑优化**和**物理优化**。**逻辑优化主要依据关系代数的等价变换做一些逻辑变换，物理优化主要根据数据读取、表连接方式、表连接顺序、排序等技术对查询进行优化**。“查询重写规则”属于逻辑优化方式，运用了关系代数和启发式规则；“查询算法优化”属于物理优化方式，运用了基于代价估算的多表连接算法求解最小花费的技术。

## 代价模型

总代价模型：COST = CPU Cost + IO Cost  
 MySQL在cost类型上分为IO、CPU和Memory，MySQL5.7的代价模型还在完善中，Memory的代价虽然已经收集了，但还没有计算在最终的代价中。  
 MySQL5.7在源码上对cost模型进行了大量重构，代价分为server层和engine层。server层主要是CPU代价，而engine层主要是IO代价。MySQL5.7 引入了两个系统表mysql.server\_cost和mysql.engine\_cost来分别配置这两个层的代价。  
以下分析均基于MySQL5.7.10

### server\_cost

1、row\_evaluate\_cost (default 0.2) 计算符合条件的行的代价，行数越多，此项代价越大；  
 2、memory\_temptable\_create\_cost (default 2.0) 内存临时表的创建代价；  
 3、memory\_temptable\_row\_cost (default 0.2) 内存临时表的行代价；  
 4、key\_compare\_cost (default 0.1) 键比较的代价，例如排序；  
 5、disk\_temptable\_create\_cost (default 40.0) 内部myisam或innodb临时表的创建代价；  
 6. disk\_temptable\_row\_cost (default 1.0) 内部myisam或innodb临时表的行代价；  
 可以看出创建临时表的代价是很高的，尤其是内部的myisam或innodb临时表。

### engine\_cost

1. io\_block\_read\_cost (default 1.0) 从磁盘读数据的cost，对innodb来说，表示从磁盘读一个page的cost；  
    2、memory\_block\_read\_cost (default 1.0）；  
    从内存读数据的cost，对innodb来说，表示从buffer pool读一个page的cost。  
   目前io\_block\_read\_cost和memory\_block\_read\_cost默认值均为1，实际生产中建议酌情调大memory\_block\_read\_cost，特别是对普通硬盘的场景。

注：分布式数据库需要增加网络开销。

MySQL8.0对于代价模型有一定的改进。

## 数据库调优

### 方式

在数据库层面进行调优，有很多的资源、数据库配置参数需要考虑。数据库调优的方式通常有如下几种：

❏人工调优。主要依赖于人，效率低下；要求操作者完全理解常识所依赖的原理，还需要对应用、数据库管理系统、操作系统以及硬件有广泛而深刻的理解。

❏基于案例的调优。总结典型应用案例情况中数据库参数的推荐配置值、数据逻辑层设计等情况，从而为用户的调优工作提供一定的参考和借鉴。但这种方式忽略了系统的动态性和不同系统间存在的差异。

❏自调优。为数据库系统建立一个模型，根据“影响数据库系统性能效率的因素”，数据库系统自动进行参数的配置。

### 阶段

通常，数据库调优主要分为5阶段，如表1-1所示。



涉及的技术主要有以下几种：

❏应用情况的估算。应用的使用方式（把业务逻辑转换为数据库的读写分布逻辑，以读多写少或读写均衡等来区分OLAP和OLTP；应用对数据库的并发情况、并发是否可以池化等）、数据量、对数据库的压力、峰值压力等做一个预估。

❏系统选型策略。确定什么样的数据库可以适用应用需求，并确定使用开源的数据库还是商业的数据库，使用集群还是单机的系统，同时对操作系统、中间件、硬件、网络等进行选型。

❏数据模型的设计。主要根据业务逻辑，从几个角度考虑表的逻辑结构，内容如下。

○E-R模型设计：遵循E-R模型设计原理。偶尔的、适当程度的非规范化可以改善系统查询性能。

○数据逻辑分布策略：目的是减少数据请求中不必要的数据量，只返回用户需要的数据。可用的技术如分区、用E-R模型分表等（如互联网企业典型的用法，根据业务的不同，进行分库、分表等操作）。

○数据物理存储策略：目的是减少IO操作，如启用压缩技术、把索引和表数据的存储分开，不同的表数据分布在不同的表空间上，不同的表空间分布在不同的物理存储上（尤其是读写量大的表空间分布在不同的物理存储上），日志、索引和数据分布在不同的物理存储上等。

○索引：在查询频繁的对象上建立恰当的索引，使索引的正效应大于负效应（索引的维护存在消耗）。

❏SQL设计。编写正确的、查询效率高的SQL语句，依据的主要是“查询重写规则”。编写语句的过程中要注意，要有意识地保障SQL能利用到索引。

❏数据库功能的启用。数据库为提高性能提供了一些功能，可合理使用，具体如下。

○查询重用：根据实际情况进行配置，可缓存查询执行计划、查询结果等。

○数据库参数的设置：可设置合适的参数，如数据缓冲区等。

❏模型系统预运行。在备用系统上模拟实际运行环境，加大压力进行系统测试，提前发现问题。

❏系统监控与分析。在工业环境下，加强对系统的运行监控和日常的分析工作，具体如下。

○应用系统表现：收集用户对应用系统的使用意见、系统存在问题等，因为这些可能是用户在第一时间发现的。

○OS环境监控：实时监控CPU、内存、IO等，并对比实时情况与历史正常情况。

○数据库内部状况监控：一些数据库提供系统表、视图、工具等手段，向用户提供数据库运行过程中内部状况的信息，如锁的情况，这些都需要实时监控，并对比实时情况与历史正常情况。

○日志分析：在数据库的日志、操作系统的日志中找出异常事件，定位问题。

## 查询优化技术

查询优化技术是SQL层面的优化，属于局部优化，有别于“数据库调优”式的全局优化。

### 查询重用技术

查询重用是指尽可能利用先前的执行结果，以达到节约查询计算全过程的时间并减少资源消耗的目的。

目前查询重用技术主要集中在两个方面：

❏查询结果的重用。在缓存区中分配一块缓冲块，存放该SQL语句文本和最后的结果集，当遇到同样的SQL输入时，可直接把结果返回。查询结果的重用技术节约了查询计划生成时间和查询执行过程的时间，减少了查询执行全过程的资源消耗。

❏查询计划的重用。缓存一条查询语句的执行计划及其相应语法树结构。查询计划的重用技术减少了查询计划生成的时间和资源消耗。

查询重用技术有利有弊：弊端，如结果集很大会消耗很大的内存资源，同样的SQL不同用户获取的结果集可能不完全相同；益处，节约了CPU和IO消耗。在使用的过程中，趋利避害，应根据实际情况选用。

注：在分布式数据库实践中，可以根据具体的执行计划复杂程度以及结果集的大小确定是否需要缓存该执行计划和结果集。

### 查询重写规则技术

查询重写是查询语句的一种**等价转换**，即对于任何相关模式的任意状态都会产生相同的结果（相同的关系替代两个表达式中相应的关系，所得到的结果是相同的）。查询重写有两个目标：

❏将查询转换为等价的、效率更高的形式，例如将效率低的谓词转换为效率高的谓词、消除重复条件等。

❏尽量将查询重写为等价、简单且不受表顺序限制的形式，为物理查询优化阶段提供更多的选择，如视图的重写、子查询的合并转换等。

查询重写的依据，是关系代数。关系代数的等价变换规则对查询重写提供了理论上的支持。查询重写后，查询优化器可能生成多个连接路径，可以从候选者中择优。

对查询优化技术进行分类，可有以下4个角度：

❏语法级。查询语言层的优化，基于语法进行优化。

❏代数级。查询使用形式逻辑进行优化，运用关系代数的原理进行优化。

❏语义级。根据完整性约束，对查询语句进行语义理解，推知一些可优化的操作。

❏物理级。物理优化技术，基于代价估算模型，比较得出各种执行方式中代价最小的。

查询重写是基于语法级、代数级、语义级的优化，可以统一归属到逻辑优化的范畴：基于代价估算模型是物理层面的优化，是从连接路径中选择代价最小的路径的过程。

查询重写技术优化思路主要包括：

❏将过程性查询转换为描述性的查询，如视图重写。

❏将复杂的查询（如嵌套子查询、外连接、嵌套连接）尽可能转换为多表连接查询。

❏将效率低的谓词转换为等价的效率高的谓词（如等价谓词重写）。

❏利用等式和不等式的性质，简化WHERE、HAVING和ON条件。

如何改进现有查询重写规则的效率，如何发现更多更有效的重写规则，是查询优化的研究内容之一。常见的查询重写技术类型，每一类都有自己的规则，这些规则没有确定的、统一的规律，但**重写的核心一定是“等价转换”，只有等价才能转换，这是需要特别强调的**。

### 查询算法优化技术

查询优化即求解给定查询语句的高效执行计划（有的书籍称为执行方案）的过程。

查询计划，也称为查询树，它由一系列内部的操作符组成，这些操作符按一定的运算关系构成查询的一个执行方案。简单说，就是先将表A和表B连接得到中间结果，然后再和另外的表C连接得到新的中间方式，直至所有表连接完毕（连接操作就是操作符，这个示例有两个连接操作符。A连接B连接C、C连接B连接A就是两种不同的执行方案，是两个不同的执行计划，查询优化要选出最高效的一个执行方案）。

查询计划，从形式上看是一颗二叉树，树叶是每个单表对象，两个树叶的父结点是一个连接操作符（如左外连接操作符，A left-out join B）连接后的中间结果（另外还有一些其他结点如排序操作等也可以作为中间结果），这个结果是一个临时“关系”，这样直至根结点。

所以从一个查询计划看，涉及的主要“关系结点”包括：

❏单表结点。考虑单表的数据获取方式，是直接通过IO获得数据，还是通过索引获取数据，或者是通过索引定位数据的位置后再经过IO到数据块中获取数据。这是一个从物理存储到内存解析成逻辑字段的过程，即符合冯·诺依曼体系结构的要求（外存数据读入内存才能被处理）。

❏两表结点。考虑两表以何种方式连接、代价有多大、连接路径有哪些等。表示的是内存中的元组怎么进行元组间的连接。此时，元组通常已经存在于内存中，直接使用即可。这是一个完成用户语义的逻辑操作，但是只是局部操作，只涉及两个具体的关系。完成用户全部语义（用户连接的语义），需要配合多表的连接顺序的操作。不同的连接算法导致的连接效率不同，如数据多时可使用Hash连接，外表数据量小且内表数据量大时可使用嵌套连接，数据如果有序可使用归并连接或先排序后使用归并连接等。

❏多表中间结点。考虑多表连接顺序如何构成代价最少的“执行计划”。决定是AB先连接还是BC先连接，这是一个比较花费大小的运算。如果判断的连接方式太多，也会导致效率问题。多个关系采用不同次序进行连接，花费的CPU资源、内存资源差异可能较大。许多数据库采用左深树、右深树、紧密树3种方式或其中一部分对多表进行连接，得到多种连接路径。

查询优化目的就是生成最好的查询计划。生成最好的查询计划的策略通常有两个：

❏基于规则优化。根据经验或一些已经探知或被证明有效的方式，定义为“规则”（如根据关系代数得知的规则、根据经验得知的规则等），用这些规则化简查询计划生成过程中符合可被化简的操作，使用启发式规则排除一些明显不好的存取路径，这就是基于规则的优化。

❏基于代价优化。根据一个代价评估模型，在生成查询计划的过程中，计算每条存取路径（存取路径主要包括上述3个“关系结点”）的花费，然后选择代价最小的作为子路径，这样直至所有表连接完毕得到一个完整的路径。主流数据库都采用了基于代价策略进行优化的技术。

基于规则优化具有操作简单且能快速确定连接方式的优点，但这种方法只是排除了一部分不好的可能，所以得到的结果未必是最好的；基于代价优化是对各种可能的情况进行量化比较，从而可以得到花费最小的情况，但如果组合情况比较多则花费的判断时间就会很多；查询优化器的实现，多是两种优化策略组合使用，如MySQL和PostgreSQL就采取了基于规则和代价估算的查询优化策略。

多表连接的优化算法中，使用最广泛的算法有如下几种：

❏SYSTEM-R算法。近乎穷举的搜索算法（一种空间搜索算法，其变形算法与其本质相同）。

❏启发式搜索算法。基于规则（基于“启发式规则”抛弃不好的存取路径挑选好的）。

❏贪婪算法。根据某种优化方式，以当前情况为基础做出最优选择，认为每次搜索过的局部存取路径是最优的，然后继续探索与其他表的连接路径。

❏动态规划算法。将待求解的问题分解为若干个子问题（阶段），按顺序求解子问题，前一子问题的解，为后一子问题的求解提供了有用的信息。在求解任一子问题时，列出各种可能的局部解，通过决策保留那些可能达到最优的局部解，丢弃其他局部解。依次解决各子问题，最后一个子问题就是初始问题的解。

❏遗传算法。一种启发式的优化算法，抛弃了传统的搜索方式，模拟自然界生物进化过程，基于自然群体遗传演化机制，采用人工进化的方式对目标空间进行随机化搜索。

我们知道，查询的基本操作是选择、投影和连接。选择和投影的优化规则适用于SPJ（Select-Project-Join）和非SPJ（SPJ+GROUPBY等操作）；连接包括两表连接和多表连接，多表连接是其中最难的，因为多个表连接时可以有多种不同的连接次序，所以查询的执行计划的数目会随着该查询包含的表个数呈指数级增长（最大组合次数是n个关系全排列），当表个数很多时，将导致搜索空间极度膨胀，仅搜索花费最小的查询计划就需要耗费巨大的时间和资源，这是查询优化器实现时需要考虑的问题。

### 并行查询的优化技术

传统单机数据库系统中，给定一个查询（Query），查询优化算法只需找到查询的一个具有最小执行花费的执行计划，这样的计划必定具有最快的响应时间。

在并行数据库系统中，查询优化的目标是寻找具有最小响应时间的查询执行计划，这需要把查询工作分解为一些可以并行运行的子工作。一些商业数据库提供了并行查询的功能，用以优化查询执行操作。

一个查询能否并行执行，取决于以下因素：

❏系统中的可用资源（如内存、高速缓存中的数据量等）。

❏CPU的数目。

❏运算中的特定代数运算符。如A、B、C、D4个表进行连接，每个表的单表扫描可以并行进行；在生成4个表连接的查询计划过程中，可选择A和B连接的同时C和D进行连接，这样连接操作能并行运行。不同商业数据库，对查询并行的实现也不尽相同。在同一个SQL内，查询并行可以分为以下两种：

○操作内并行。将同一操作如单表扫描操作、两表连接操作、排序操作等分解成多个独立的子操作，由不同的CPU同时执行。

○操作间并行。一条SQL查询语句可以分解成多个子操作，由多个CPU执行。

注：分布式数据库实践中，对于多表UNION操作可以采用拆分多个SQL，并行下发执行的优化方式提高效率。

### 分布式查询优化技术

在分布式数据库系统中，查询策略优化（主要是数据传输策略，A、B两结点的数据进行连接，是A结点数据传输到B结点或从B到A或先各自进行过滤然后再传输等）和局部处理优化（传统的单结点数据库的查询优化技术）是查询优化的重点。

注：在分布式数据库中不同节点之间数据传输会耗费大量的网络IO，这个是重点考虑优化的地方。

在查询优化策略中，数据的通信开销是优化算法考虑的主要因素。分布式查询优化以**减少传输的次数和数据量**作为查询优化的目标。所以，分布式数据库系统中的代价估算模型，除了考虑CPU代价和IO代价外，还要考虑通过网络在结点间传输数据的代价。这是分布式并行查询优化技术与传统单结点数据库系统最大的不同之处。

在分布式数据库系统中，代价估算模型如下：

总代价=IO代价+CPU代价+通信代价

### 其他优化技术

数据库的查询性能，还取决于其他一些因素，如数据库集群系统中的SD（Share Disk）集群和SN（Share Nothing）集群，不同的架构查询优化技术也不同。SD集群采用的是共享存储方式，在数据的读写时可能产生读写冲突，所以单表扫描会受到影响；SN集群采用的是非共享式存储方式，所以在考虑了通信代价后单结点的优化方式依然适用。

# 逻辑查询优化

查询优化器在逻辑优化阶段主要解决的问题是：如何找出SQL语句等价的变换形式，使得SQL执行更高效。

一条SQL查询语句结构复杂，包含多种类型的子句，优化操作依赖于表的一些属性信息（如索引和约束等）。可用于优化的思路包括：

❏子句局部优化。每种类型子句都可能存在优化方式，这是子句局部的优化，如等价谓词重写、WHERE和HAVING条件化简中的大部分情况，都属于这种子句范围内的局部优化。

❏子句间关联优化。子句与子句之间关联的语义存在优化的可能，如外连接消除、连接消除、子查询优化、视图重写等都属于子句间的关联优化，因为它们的优化都需要借助其他子句、表定义或列属性等信息进行。

❏局部与整体的优化。需要协同考虑局部表达式和整体的关系，如OR重写并集规则需要考虑UNION操作（UNION是变换后的整体的形式）的花费和OR操作（OR是局部表达式）的花费。

❏形式变化优化。多个子句存在嵌套，可以通过形式的变化完成优化，如嵌套连接消除。

❏语义优化。根据完整性约束、SQL表达的含义等信息对语句进行语义优化。

❏其他优化。根据一些规则对非SPJ做的其他优化、根据硬件环境进行的并行查询优化等。各种逻辑优化技术依据关系代数和启发式规则进行。

## 关系代数

## 查询重写规则

传统的联机事务处理（On-line Transaction Processing，OLTP）使用基于选择（SELECT）、投影（PROJECT）、连接（JOIN）3种基本操作相结合的查询，这种查询称为SPJ查询。

数据库在查询优化的过程中，会对这3种基本操作进行优化。优化的方式如下：

❏选择操作。对应的是限制条件（格式类似field＜op＞consant，field表示列对象，op是操作符，如=、＞等），优化方式是选择操作下推，目的是尽量减少连接操作前的元组数，使得中间临时关系尽量少（元组数少，连接得到的元组数就少），这样可减少IO和CPU的消耗，节约内存空间。

❏投影操作。对应的SELECT查询的目的列对象，优化方式是投影操作下推，目的是尽量减少连接操作前的列数，使得中间临时关系尽量小（特别注意差别：选择操作是使元组的个数“尽量少”，投影操作是使一条元组“尽量小”），这样虽然不能减少IO（多数数据库存储方式是行存储，元组是读取的最基本单位，所以要想操作列则必须读取一行数据），但可以各减少连接后的中间关系的元组大小，节约内存空间。

❏连接操作。对应的是连接条件（格式类似field\_1＜op＞field\_2，field\_1和field\_2表示不同表上的列对象，op是操作符，如=、＞等），表示两个表连接的条件。这里涉及以下两个子问题。

○多表连接中每个表被连接的顺序决定着效率。如果一个查询语句只有一个表，则这样的语句很简单；但如果有多个表，则会涉及表之间以什么样的顺序连接效率最高效（如A、B、C三表连接，如果ABC、ACB、BCA等连接后的结果集一样，则计算哪种连接次序的效率最高，是需要考虑的问题）。

○多表连接每个表被连接的顺序由用户语义决定。查询语句多表连接有着不同的语义（如是笛卡儿集、内连接，还是外连接中的左外连接等），这决定着表之间的前后连接次序是不能随意更换的，否则，结果集中数据是不同的。因此，表的前后连接次序是不能随意交换的。

另外，根据SQL语句的形式特点，还可以做如下区分：

❏针对SPJ的查询优化。基于选择、投影、连接3种基本操作相结合的查询。

❏针对非SPJ的查询优化。在SPJ的基础上存在GROUPBY操作的查询，这是一种较为复杂的查询。

所以，针对SPJ和非SPJ的查询优化，其实是对以上多种操作的优化。“选择”和“投影”操作，可以在关系代数规则的指导下进行优化。表连接，需要多表连接的相关算法完成优化。其他操作的优化多是基于索引和代价估算完成的。

### 子查询优化

### 视图重写

### 等价谓词重写

### 条件化简

### 外连接消除

### 嵌套连接消除

### 连接消除

### 语义优化

### 针对非SPJ优化

## 启发式规则在逻辑优化阶段的应用

# 物理查询优化

## 查询代价估算

## 单表扫描算法

## 索引

## 两表连接算法

## 多表连接算法

# 查询优化器原理解析

## 概述

MySQL查询优化器主要的功能是完成SELECT语句的执行，在保证SELECT语句正确执行之外，还有一个重要的功能，就是使用关系代数、启发式规则、代价估算模型等不同种类的技术，提高SELECT语句执行的效率。

### 查询执行过程

MySQL查询执行过程分为4个阶段，如下所示。

❏语法分析阶段：将SQL查询语句经词法和语法分析后变换为一棵查询树st\_select\_lex传给优化器，并对SQL表达的语义进行检查。

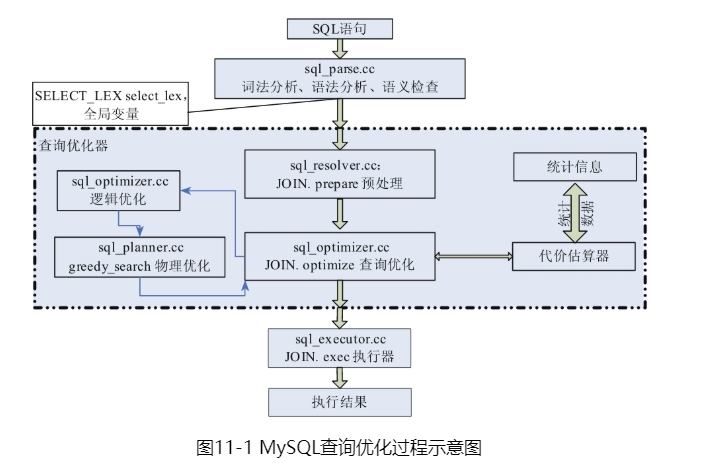
❏生成逻辑查询执行计划阶段：优化器在查询树中遍历每个关系，确定关系是否是常量表、为每个关系查找可用的索引、运用关系代数原理和启发式规则进行逻辑上的查询优化（如消除子查询、消除外连接等）。

❏生成物理查询执行计划阶段：优化器对各个连接的表进行排序，然后再求解多表连接最优路径，对于每个关系尽量利用索引计算其代价，找出代价最小的路径后保存到JOIN类的best\_positions。

❏执行查询执行计划阶段：把查询执行计划传到执行器进行执行。4个阶段如图11-1所示。

MySQL查询优化器在逻辑查询执行计划阶段，基于关系代数规则和启发式规则，把用户指定的SQL经过“等价”的代数转换，变为一种更节省IO的执行序列，执行起来更为高效。

MySQL查询优化器在物理查询执行计划阶段，在解决多表连接的问题时，有两套算法：一是用户指定表连接次序的算法；二是混杂了贪婪和穷举思想的算法，解决的是较多表的连接或非用户指定连接次序的多表连接，但不能保证得到最优的查询执行计划。



## 数据结构

## 原理

## 查询优化

## 关键算法