

并行处理提高 Oracle 工作性能的研究

Study on Parallelism Improving the Performance of Oracle

武欣嵘 (解放军理工大学通信工程学院电信工程系 江苏南京 210007)

袁晓勇 (解放军理工大学通信工程学院研究生三队 江苏南京 210007)

摘要:数据库的工作性能直接决定应用系统的工作性能。并行处理可以提高 Oracle 数据库的工作性能。本文在对 Oracle 并行原理、并行参数设置的基础上,对 Oracle 并行处理功能的使用方法、使用条件提出了自己的观点,为用户合理使用并行处理提供了方便

关键词:操作内并行 操作间并行 并行度 并行 SQL

1 引言

系统的响应时间是计算机应用系统非常关心的一个系统指标。在以数据库为基础和核心的计算机应用系统中,数据库的工作性能对系统的响应时间具有关键的作用。当前,各种中大型应用系统中应用最广泛的数据库系统是 Oracle 数据库系统,因而如何提高 Oracle 的工作性能成了值得关注的问题。除了升级硬件、书写高质量的 SQL 语句外,良好的数据库设计也是提高 Oracle 工作性能的另一重要途径。从 Oracle8i 开始,企业版的 Oracle 数据库提供了并行处理的功能选项。该选项若使用恰当会大大地提高数据库的工作性能;但如果该选项使用不当,收效甚微,甚至会降低工作性能。如何利用 Oracle 提供的并行性功能选项,进行良好的数据库设计,提高数据库的工作性能,成了一个很值得数据库设计者研究的问题。

2 Oracle9i 并行性原理

SQL (Structured Query Language, 结构化查询语言) 是用户访问和操作 Oracle 数据库的唯一途径。Oracle 数据库在接受到 SQL 命令后,会把 SQL 命令分解为一系列操作。Oracle 的并行是在操作级别上进行的,包括操作内并行和操作间并行。

2.1 操作内并行

在 SQL 语句 `select * from dept` 中包含一个扫描表 dept 的操作。如果串行地执行该操作时,一个服务器“工作进程”完成所有的操作,工作过程如图 1 所示。

并行地执行该操作时 Oracle 会使用多个服务器“工作进程”。其中,一个服务器“工作进程”,作为“并行调度器”,对操作数据进行分解,然后分发给各个作为并行“工作进程”的服务器“工作进程”,并接受来自各个并行“工作进程”返回的结果,最后对这些结果进行组织并把总的结果返回给用户。

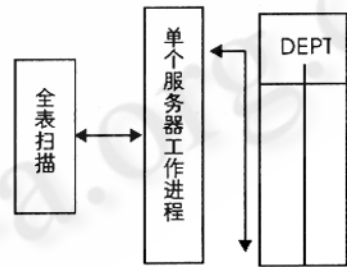


图 1 串行操作

图 2 描述了该操作被并行处理的情况。“并行调度器”把存储表 dept 数据的“数据块”(Data Block),分成一个个的工作单元,称为工作“粒度”。每个工作“粒度”由一系列连续的“数据块”(Data Block)构成。每个工作“粒度”只能由“并行调度器”分配给一个并行“工作进程”。每个工作“粒度”在执行时由“并行调度器”随机分配给并行“工作进程”的。并行“工作进程”把所分配的工作“粒度”执行完毕后,“并行操作调度器”会重新给它分配一个新的工作“粒度”,直至所有的工作“粒度”被执行完毕。所有的工作“粒度”被执行完毕后,“并行调度器”对并行“工作进程”返回的结果进行组合,把总的结果返回给用户。

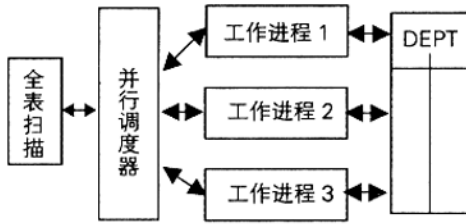


图 2 操作内并行

显然, Oracle 并行执行一个操作时, 要比串行处理该操作的速度要快一些。这也就是在硬件资源满足的情况下, Oracle9i 并行操作选项有助于提高数据库的工作性能的原因。

2.2 操作间并行

对于执行计划中只包含多个操作的并行 SQL 命令, Oracle 还需要操作间并行。为了说明 Oracle 是如何实现操作间的并行, 现以如下 SQL 语句为例进行说明:

```
select * from dept order by deptno.
```

在上面 SQL 命令的执行计划中包含两个操作: 一个操作是对库表 dept 进行扫描, 记为 Operate1; 另一个操作是基于列 deptno 对所得的结果集进行排序, 记为 Operate2; Operate2 的数据来自 Operate1。如图 3

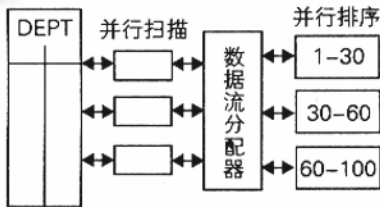


图 3 操作间并行

所示, 并行执行该 SQL 命令时, 需要在这两个操作间进行并行。在这两个操作间进行并行时, Oracle 会为这两个操作分别分配一个服务器“工作进程”组。即 Operate1 和 Operate2 各有一个服务器“工作进程”组。两个操作间有一个“数据流分配器”。“数据流分配器”的作用是把来自并行扫描的“工作进程”的数据, 正确地分配给负责并行排序的“工作进程”。在图 3 中, 在负责并行排序的“工作进程”中, 进程 1 负责 deptno 值处于范围 1~30 的记录的排序; 进程 2 负责 deptno 值处于范围 30~60 的结果集的排序; 进程 3 负责 deptno 值处于范围 60~100 的结果集的排序。如果扫描到一条记录, 其 deptno 值处于 1~30 中, 该记录就会被“数据流分配器”分配给进程 1 处理; 其 dept-

no 值处于 30~60 中, 该记录就会被“数据流分配器”分配给进程 2 处理; 其 deptno 值处于 60~100 中, 该记录就会被“数据流分配器”分配给进程 3 处理。这样通过“数据流分配器”可以成功地实现负责并行扫描的“工作进程”和负责并行排序的“工作进程”同时工作, 实现了这两个操作间的并行。

3 并行处理参数设置

要使用 Oracle 的并行处理功能, 需要进行一些必要的参数设置后。这些参数设置包括: 参数调整模式的选择; 继承参数的设置; 操作并行度的设置。

3.1 参数调整模式

参数调整模式有: 手工调整和自动调整。自动调整时, Oracle 根据硬件系统中 CPU 的数目、参数 PARALLEL_THREADS_PER_CPU 的值、系统的负载自动决定并行处理的各种参数设置。一般情况下, 采用自动调整模式时, Oracle 给出的参数可使数据库的工作性能处于最佳。也可以手动的调整并行处理参数。手动调整地缺点是: 系统的负载和资源使用情况难以做出正确估计, 也就很难给出最佳参数配置; 需要时刻根据系统负载的变化对参数进行调整。故一般推荐自动调整参数。只需把参数 PARALLEL_AUTOMATIC_TUNING 设置为 TRUE, 即可实现自动调整并行处理参数。

3.2 继承参数的设置

PARALLEL_AUTOMATIC_TUNING 设置为 TRUE, 有一些参数会被自动设置, 这些参数被称为“继承参数”。“继承参数”包括: PARALLEL_ADAPTIVE_MULTI_USER (简称参数 1); PARALLEL_MAX_SERVERS (以下简称参数 2)。“继承参数”在设置前后的值如表 1 所示。如果用户认为被 Oracle 自动设置的“继承参数”不能使性能最佳, 可以对这些参数进行手动的设置以达到最佳工作性能。

表 1 继承参数变化情况

继承参数	设置前	设置后
参数 1	false	True
参数 2	5	CPUs x 10

3.3 并行度 (Degree of Parallelism, 简称 DOP) 的设置

操作的“并行度”(DOP)是指“并行调度器”分配给该操作的并行“工作进程”数目。“并行度”(DOP)的设置有以下几种方式:

(1) 命令级设置。在 SQL 命令中利用 parallel 提示(hint)在命令级完成“并行度”(DOP)的设置。如:

```
select /* +parallel(emp,4) */ count(*) from emp
```

以上 hint 提示使扫描表 emp 的并行度是 4。

(2) 会话级设置。利用命令 ALTER SESSION FORCE PARALLEL <N>, 在会话级完成“并行度”(DOP)的设置,其中 N 是并行度。

(3) 表级或索引级设置。并行是表或索引的属性选项,可以在创建表或索引时对该选项进行说明,确定并行度。如果在创建表或索引时未对该属性进行说明,可以用 ALTER 命令进行并行度设置。如:

```
ALTER TABLE emp PARALLEL 4;
```

以上命令把在表 emp 上进行并行操作的并行度均设置为 4。

4 并行 SQL 语句

在并行操作参数设置后,并行 SQL 是用户使用 Oracle 并行操作功能的专有工具。并行化 SQL 可以被分为:并行 SELECT、并行 DML、并行 DDL。其中,DML 语句包括 INSERT 语句、UPDATE 语句、DELETE 语句;DDL 语句包括 CREATE INDEX 语句、CREATE TABLE...AS SELECT 语句、ALTER INDEX...REBUILD 语句。另外,DML 语句和 DDL 语句通常包含有查询操作,在下面讨论的 DML 语句和 DDL 语句的并行性和并行度时专指 DML 操作和 DDL 操作,不包括查询操作。一条 SQL 要被 Oracle 并行执行需要确定其并行性及并行度。不同类型的 SQL,其确定并行性和确定并行度规则是各不相同的。

4.1 并行 SELECT

SELECT 语句并行性的确定须满足下列条件:

(1) SELECT 语句加入 Parallel 提示(hint)或者其操作的表或索引被设置 Parallel 属性;

(2) SELECT 语句执行计划中至少包含一个可并行化执行的操作。

SELECT 并行度的确定原则:命令级设置 > 表级或索引级设置

4.2 并行 DML 语句

DML 语句只有在分区表上才具有并行性。除此之外,DML 语句并行性的确定,还须满足下列条件:

(1) DML 语句执行前,执行了命令 ALTER SESSION FORCE PARALLEL DML

(2) DML 操作对应得库表具有 Parallel 属性或 DML 语句带有 PARALLEL 提示(hint)。

DML 并行度的确定原则是:命令级设置 > 在会话级设置 > 表级或索引级设置。但是 DML 并行度最大不会超过表的分区数。

4.3 并行 DDL 语句

DDL 语句并行性的确定有两种方式:

(1) 在命令级,DDL 语句中加入子句 Parallel;

(2) 在会话级,执行命令 ALTER SESSION FORCE PARALLEL DDL。

DDL 并行度的确定原则是:命令级 > 会话级。

5 合理使用并行处理

并行处理大数据量操作往往能极大提高其工作性能。这些大数据量的操作通常有:

(1) 大表的扫描(Scan)和连接(Join);

(2) 大索引的创建;

(3) 分区索引的扫描;

(4) 大批数据的的插入(INSERT)、修改(UPDATE)、删除(DELETE);

(5) 大数据量的聚类(Group by)和排序(Order By)。

在数据决策系统(Decision Support System)和数据仓库(Data Warehouse)中经常会有这样的大数据量操作。因此,在数据决策系统(Decision Support System)和数据仓库(Data Warehouse)中常常需要用到并行操作。

数据库系统所在服务器的硬件处理能力也是决定是否采用并行处理的关键因素。具有下列特性的硬件系统更适合进行并行处理:

(1) 具有足够高的 I/O 带宽;

(2) CPU 利用率一直性地或间断性较低;

(3) 具有充足的内存。

一个应用如果大数据量的操作较少或者系统硬件资源不足,建议不要使用并行处理。在这种情况下,使

(下转第 118 页)

(上接第 114 页)

用并行处理收效甚微甚至可能降低整体性能。例如某个数据仓库系统具有 64 个 CPU, 如果有 500 个并发用户, CPU 较忙, 就不适采用合用并行理; 如果有 5 个并发用户, CPU 将会有许多的空闲时间, 适合采用并行处理。

6 结束语

本文对 Oracle 数据库并行性的原理, 参数的设置, 并行 SQL 的使用进行了详细的说明, 并对合理使用并行处理提出了一定的建议。并行处理是 Oracle 数据库的一大优点, 合理地使用它, 会提高数据库工作性能, 降低应用系统时间。在使用 Oracle 数据库的并行性时, 一定要针对数据库系统的实际情况进行分析, 要正确判断是否需要并行处理, 尤其是对于硬件资源不足的系统使用 Oracle 的并行性会加重系统的负担反而降低工作性能。只有根据实际正确分析, 才能真正用好并行性, 提高系统的潜在处理能力。

参考文献

- 1 THOMAS KYTE . Oracle 高效设计[M], 钟鸣、郝玉洁等译, 机械工业出版社, 2006.
- 2 盖国强、冯春培等, Oracle 数据库性能优化, 人民邮电出版社, 2005.
- 3 MIXHELE CYRAN. Oracle Database Concepts, 10g Release 2.
http://download-west.oracle.com/docs/cd/B19306_01/server.102/b14220.pdf. 2005.
- 4 PAUL LANE . Oracle Database Data Warehousing Guide, 10g Release 2.
http://download-west.oracle.com/docs/cd/B19306_01/server.102/b14223.pdf. 2005.
- 5 . STEVE FOGEL. Oracle Database Administrator's Guide, 10g Release 2.
http://download-west.oracle.com/docs/cd/B19306_01/server.102/b14231.pdf. 2006.