

1 本节我们讨论结构化数据的大规模并发处理模型

2 数据处理系统提供大数据计算处理能力和应用开发平台。

从计算架构的角度，将数据处理系统分为算法层、计算模型层、计算平台和引擎层等。

与大数据相关的计算算法包括机器学习算法和数据挖掘算法。

计算模型是指不同类型的大数据在不同场景下的处理方式，

包括批处理、流计算、结构化数据的大规模并发处理(MPP)模型、内存计算模型和数据流图模型。

就计算平台和引擎而言，通常具有代表性的大数据处理平台有 Hadoop、Spark、storm、Pregel 等，本节我们讨论结构化数据的大规模并发处理模型 MPP

3 大规模并行处理（MPP，massively parallel processing）是多个处理器（processor）处理同一程序的不同部分时该程序的协调过程，工作的各处理器运用自身的操作系统（Operating System）和内存。

大规模并行处理器一般运用通讯接口交流。在一些执行过程中，高达两百甚至更多的处理器为同一应用程序工作。数据通路的互连设置允许各处理器相互传递信息。

一般来说，大规模并行处理（MPP）的建设很复杂，这需要掌握在各处理器间区分共同数据库和给各数据库分派工作的方法。大规模并行处理系统也叫做“松散耦合”或“无共享”系统。

一般认为，对于允许平行搜索大量数据库的应用程序，大规模并行处理（MPP，massively parallel processing）系统比对称式并行处理系统（SMP）更好。可以更好支持决策支持系统（decision support system）和数据仓库（data warehouse）应用程序。

4 大数据可能是半结构化的，也可能是非结构化的。

大规模并行处理(MPP)体系结构构建了大数据，以方便报告和分析目的的查询。

MPP 系统有时被称为无共享系统。

这意味着数据跨多个服务器(也称为节点)进行分区，每个服务器在本地处理查询。

在图中，使用 4 个节点替换单个服务器，将查询时间从 1 小时减少到 15 分钟。

5 MPP 处理流程

该过程由客户端发出查询开始，然后将查询传递给主节点。

主节点包含信息，如数据字典和会话信息，

它用来生成一个执行计划，旨在从每个工作节点检索所需的信息。

并行执行是指通过节点 1 到节点 n 的并行计算来实现执行计划。 查询结果返回到主节点

6 大规模并行处理是指多个处理器协调处理单个任务，

每个处理器使用自己的操作系统和内存，并使用某种形式的消息传递接口彼此通信

MPP 可以使用无共享架构或共享磁盘架构进行设置

在无共享体系结构中，系统中不存在单点争用，节点不共享内存或磁盘存储。

数据在节点之间水平分区，这样每个节点都有一个来自 DB 中每个表的子行集

然后，每个节点只处理自己磁盘上的行。

具有分布式数据库的“无共享”系统需要大量的协调才能完成一项公共任务。

每个节点都拥有数据库的片段。 管理这个数据库可能非常困难。具有副本数据库的无共享系统不适合具有大量数据需求的应用程序。

如果计算需要大量数据修改操作，如数据插入和连接，那么“无共享”体系结构可能是不可用的。

7 通过 Segment 实例并行性实现的性能

架构有 Master host and standby master Host

主主机和备用主主机

Master Host 与 Segment Host 分段主机协调工作

Segment Host 主机与一个或多个 Segment 实例一起工作。Segment 实例并行处理查询

Segment 主机有自己的 CPU 磁盘内存(不共享任何内容)

高速互连用于持续数据处理的 Pipeline 流水线

8 MPP 数据库体系结构

MPP 数据库体系结构。有一个主服务器和 4 个从服务器，每个从服务器上有两个 segment，每个 segment 在另一台机器上有一个镜像。

MPP 数据库架构：典型的 MPP 数据库通常采用无共享架构，由一个主节点和 n 个从节点组成。主节点负责与客户端交互、管理整个集群并协调查询处理。

n 个辅助节点中的每个节点负责存储数据的一个分区，并在其分区上执行查询处理。每个从节点承载 d 个数据库实例，后面将它们称为 segment。

大多数 MPP 数据库在存储级别提供容错功能。镜像方案，即复制，通常用于确保数据的持久性和可用性。

如图所示，每个 segment (主段)在另一个节点上分配一个镜像(镜像段)。

主节点通过监控从节点的心跳来检测节点故障。如果从节点停止响应一段时间，称为系统延迟时间(通常在 1 分钟左右)，

主节点将把它视为失败的节点。一旦检测到故障，相应的镜像将被激活以替换故障的主 segment。

类似地，备用节点充当主节点的复制/镜像。通过镜像方案，可以显著提高系统的可用性。

但是，这种镜像方案不支持自动的查询内部容错。

当节点发生故障时，故障节点上正在运行的查询的状态将丢失。

激活相应的镜像后，必须重新运行整个查询。

如果它是一个长时间运行的查询，对客户端的响应将严重延迟。在最坏的情况下，如果失败的概率很高查询将无限期地运行。

9 Greenplum

在大规模并行处理(MPP)数据库中，数据被划分到多个服务器或节点上，每个服务器/节点都有内存/处理器来本地处理数据。所有的通信都是通过网络互连的-没有磁盘级的共享或要关注的争用(即它是一个“无共享”的体系结构)。

1)主主机—独立的物理服务器，拥有自己的操作系统/CPU/存储/内存。

2) 主机主数据库。主数据库中存储用户数据，但存储了关于 segment 的元数据——从系统表的角度考虑。

3) 2、3、4segment 主机——拥有自己的操作系统/CPU/存储/内存的独立物理服务器。

4) 主机 segment 数据库。每个数据库都存储部分用户数据。

5) 互连交换机——segment 服务器数据库通过互连交换机进行通信

MPP 数据库的主要特点是数据分布。

数据分布在各个分段数据库中，以实现数据处理的并行性。

这是通过使用 DISTRIBUTED by 子句创建数据库表来实现的。

通过使用该子句，数据自动分布在各个段数据库中。

在 Greenplum 中，您可以使用散列分布或循环分布。

10 本节我们讨论了结构化数据的大规模并发处理模型，今天的内容就到这里，谢谢大家