

分布式数据库面临的问题及未来发展趋势

张娜

(北京师范大学研究生院珠海分院 广东珠海 519000)

摘要： 在云计算、大数据等新兴技术的推动之下，现有的单机存储已经无法满足目前海量数据的存储和高效计算的需要。迫使企业对数据库的需求指向大规模、高可靠、高扩展及高性能，分布式数据库应运而生。但是随着业务更加复杂化、多样化，对于数据也将更加错综复杂，对数据库的要求也会日渐提高。当前分布式数据库仍然有问题亟待解决，本文将就当前分布式数据库所面临的问题以及未来发展的潜力及趋势进行讨论。

关键词： 分布式数据库；海量数据；数据库水平拓展；

1 引言

数据库的发展可以从 1970 年代开始说起，数据库最基本的功能有 2 个，一是将数据完整的保存下来；二是满足用户对数据的计算需要。第一点是最基本的要求，如果一个数据库无法保证所有数据的完整安全的保存，则后续的操作将变得毫无意义；在满足第一点的前提下，还要满足用户对数据的相关操作，可能是简单的查询也可能是复杂的计算，如聚合操作，连表查询等等。

在数据库发展的早期，这两点基本要求不难达成。但是，随着大数据、云计算不断的蓬勃兴起，而这些新兴技术均依赖于大量数据整合之后的管理和分析。面对这些需要成百上千的 GB 的字节数据量，甚至是 PB 级别如此大体量的巨型存储文件，目前单机存储已经无法满足基本的存储需要，更无从谈起用户计算需要。此外，对于多级部署方式的数据库，越靠近顶端所需承载的数据量也会变大，数据的爆炸性增长，数据在线时间不断延长。互联网的蓬勃兴起，使得目前用户数也呈现爆炸性增长，用户数增多，相应应用的使用请求数也会相应增长，即使开发者通过软件层面，对用户应用的请求进行优化，但对于降低数据库压力也如同杯水车薪。这种种情况综合之后，带来一系列的问题，如数据创建缓慢，管理低效，读写耗时，备份困难等等诸多问题，分布式数据库这种解决方案便应运而生。

2 分布式数据库简介

广义上，分布式数据库是用计算机网络将物理上分散多个数据库单元链接起来，组成一个逻辑上统一的数据库。每个被链接的数据库单元称为站点或者节点，每一个节点都是一个独立的数据库管理系统。分布式数据库管理系统可以视作是一系列的数据库管理系统的集合，管理着多个数据库系统。但是分布式数据库管理系统并非是一个简单的集合。如果谈及分布式数据库，我们不可避免要谈及另外一个概念，“数据库集群”。数据库集群和分布式数据库相似，也是用计算机网络将物理上分散多个数据库单元链接起来，组成一个逻辑上统一的数据库。

数据库集群和分布式数据库虽然都是由多个分散开来的数据库单元链接起来，但是其中区别是非常大的。

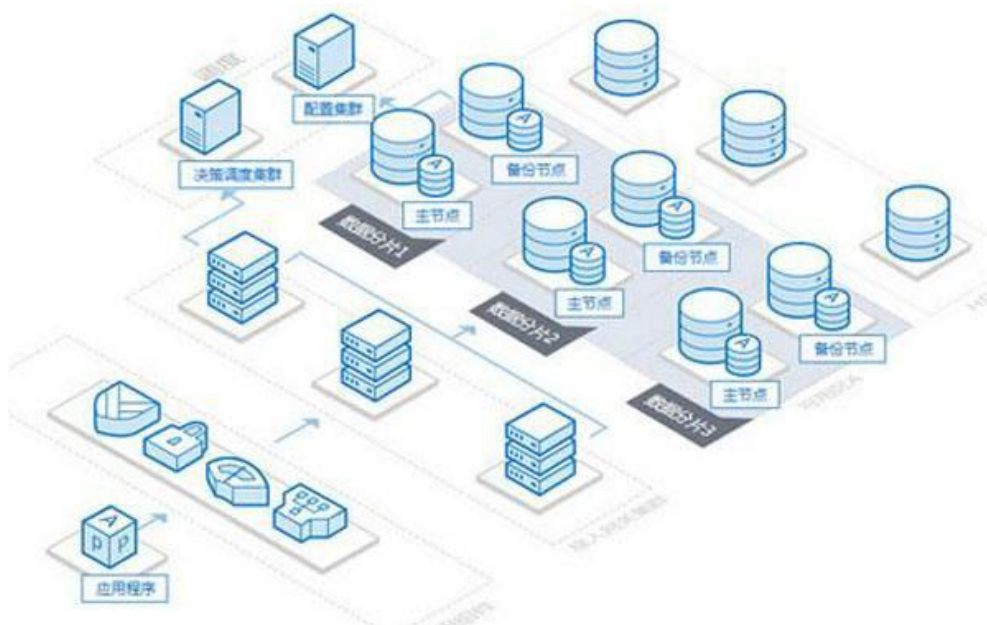


图 1 分布式数据库结构图

数据库集群的每一个数据库单元具有单份数据集，有的甚至有多份相似的数据集，且每个数据单元往往存在于高速的局域网内，而分布式数据库则往往拥有完全不同的数据集，分布式数据库是将一整个数据集切片分区，分散在不同的数据节点之中，每个数据单元可以在高速局域网内，亦可以是在不同的片区中，阿里巴巴的中美异地机房分布式数据库系统便是一个很好的例子。分布式数据库结构如图 1 所示。有一个中央的根主机，另外还有若干个数据节点以及其备份节点，每个数据节点都是一个相对独立的数据库管理系统，实现“局部自治”。此外，

数据库集群往往为同构的系统，要求每一个节点都具有相同的操作系统及数据库系统版本，而分布式数据库则没有此要求，可以是异构的系统，即是可以有不同的操作系统，不同的数据库版本。

通过上述比较，我们可以得知，通过群集数据库的设计方式可以解决测点不断扩展等问题，但是解决不了巨型数据文件的问题。同时对于单库吞吐率的也没有帮助。数据库测点数量与历史数据文件的大小直接决定了分布式存储建设规模，特别是数据库服务器的配置数据量。假设单台服务器能够支持的测点容量为 X（测点容量与具体数据库产品有关），同时实时数据库服务器采用主备方式运行，某阶段建设时估算的所有测点数量为 Y（包括了测点容量裕度），则此阶段所需的数据库服务器数量为

$$(\text{INT}(Y/X)+1) * 2$$

例如：一个实时/历史数据库的最大测点容量为 20 万，某应用的所有测点数量为 100 万，则所需服务器的数量为

$$(\text{INT}(100/20)+1) * 2 = 12$$

还是同样的应用场景，一个 20 万点数据库，设置为不压缩，不算索引文件和卫星数据，每秒钟数据全刷新一次，所产生的数据量为

$$(8B V + 2B Q + 8B T) * 200\ 000 = 3.4MB$$

每年产生的数据量 102TB，在这样的场景之下，单服务器处理数据的能力迅速下降。可见简单的“群集数据库”不能解决数据文件不断膨胀带来的单服务器处理能力不足的问题。

而分布式数据库，通过将数据进行切块，然后通过一个统一的分布式数据库管理系统，对数据进行管理和操作则可以解决单服务器的数据存储和数据处理不足的问题。

3 分布式数据库特点及优势

分布式数据库管理系统是在集中式数据库系统的基础上发展演变而来，在物理上，不同的节点分散在不同的地区，数据库中的数据也分别放在不同的节点，即不同的数据分区，不同的局部数据去存储，由各自的数据库管理系统去管理。而这些节点可能是不同的操作系统，不同的数据库版本，但是通过网络通信链接

在一起，形成一个逻辑上统一的整体。然而，从使用者的角度来说，它们所构成的是一个完整的数据库，使用上并没有感觉到什么区别，如同集中式数据库系统一般。这就是分布式数据库系统的实现的基本目标。

3.1 分布式数据库特点

分布式数据库基本特点有如下几点：

(1) 物理分布性

在物理上，节点是分散在不同的区域，数据也是分片存储在不同区域，因为在大多数网络环境之下，单机存储最终将无法满足实际需要，分布式数据库的物理分布性可以支持水平拓展，进一步分担数据存储压力和分担处理任务的压力。

(2) 逻辑整体性

逻辑上，这些节点相互链接，形成一个逻辑上统一的整体，由于分布式数据库是依据数据分片，每个数据节点存放着不相一致的数据集，分布式数据库在逻辑上是一个整体，对于用户来说，用户无需关心数据的具体分区，各个节点具有相对的自治，可以完成各自的数据处理任务，同时系统内又包括了一个集中控制的全局管理系统，协调各局部节点的工作，执行全局任务。这样就实现了逻辑上的整体性。

(3) 站点自治性

分布式数据库采用集中和自治相结合的控制结构，每个节点都是一个独立的数据库管理系统，满足集中式数据库的一致性、可串行性和可恢复性。

3.2 分布式数据库优势

分布式数据库优势有如下几点：

(1) 可以随时针对各区域的用户进行调整，可以根据实际需要增加或减少相关的节点数量；

(2) 数据共用和分布式控制，数据库是用户共享的资源，为了保证数据库的安全和完整性，对于共享数据库的控制是集中的。在分布式数据库共享有两个层次，一个是局部共享，一个是全局共享。因此，相应的控制结构也有两种，集中和自治，不同的系统集中和自治的程度不尽相同。

(3) 增加处理绩效，可以作平行处理，每个节点实际上是相对独立的数据库管理系统，所以在处理数据时，如果属于不同区域的数据，可以实现

(4) 系统管理维护成本低，质量控制比较容易。

4 分布式数据库当前存在问题

分布式数据库是数据库发展至今适应当前发展的产物，云服务和大数据业务更加蓬勃兴起，更多会存储用户信息，用户行为信息，这就产生了非常庞杂的数据，如何实现在存储这些数据之后，性能问题就慢慢走入人们的视线，成为开发者更加关注的问题。分布式数据库可以解决存储问题，但是如果涉及到更加深入的话题，如数据库性能，如何使数据库发挥最大化作用，这就是当前分布式数据库所要面临的问题。

如双十一期间，淘宝在 2013 年创下了 79 万笔/分钟的交易峰值，2014 年的双十一，淘宝交易峰值达到 285 万/分钟，系统处理能力要求提高了近 3 倍。并且还要保持数据的强一致性和数据的零丢失的要求，这对分布式数据库的要求更加严格。

为了解决数据库在存储目标达成之后的性能需求，有如下的几个大致的目前仍然需要考虑的问题：

(1) 数据分片策略

我们都知道，分布式数据库每个数据节点往往都是不一样的数据集，解决了容量问题，还要避免分库分表对业务带来的影响，实现跨行跨表的事务。目前对数据分片的策略有对日期(Date)顺序进行分片，基于字段范围(Range)进行分片，但是这两种都有一个比较明显的不足，这样的分片策略会导致数据的严重倾斜，分片和分片之间可能会出现负载和数据容量的不均衡，因为在大部分的数据库系统之中，如淘宝热搜，微博热搜，这样子的业务会导致数据有明显的冷热的特征，某个部分的数据的访问概率会比其他数据更高，如果采用这两种分片策略，会导致热数据会存在于少数的分片之中，而存储冷数据的设备性能将会被浪费掉。

这对这种情况腾讯云采用的是某个字段通过 HASH 求模的方案进行分片，将求模后的字段范围分散到不同分片之中。HASH 算法可以基本保证数据相对均匀分散在不同的物理主机之中，当 SQL 请求发起时，按照拆分键和执行策略将 SQL

路由对应到不同的对应表进行执行。阿里团队也提出了相似的分片策略，对某个业务范围进行水平拆分，通常是对用户编号 HASH 求模，将求模结果分散到不同的物理主机，但是这种方式当涉及到复杂的操作的时候，有些范围查询会需要访问到所有主机，性能低下。另外阿里团队也提出了对业务范围垂直拆分数据库，根据业务特征将数据集中分布到不同的数据库中，但是这个适用于业务和业务之间耦合度低的情况，如果业务间的耦合度高，则拆分能力将会受限，拓展性也会受到限制，跨库事务的一致性需求强，程序的复杂度也会提高，从而影响到系统的稳定性和开发效率。

选择一个合理的数据库分片策略对于后续开发将会产生比较深远的影响。

(2) 跨行跨表的事务管理功能

实现数据跨服务器分布之后，跨行跨表事务功能变得难以处理。理论上，一个事务不会涉及多个分片数据，但是目前业务需求复杂，跨机事务变得不再是不可能。实现分布式事务处理的最大难点，就是在这么多个数据节点实例上实现数据库事务的 ACID 保障。当前的跨机事务常见的办法有利用两阶段提交支持分布式事务，腾讯云也应用了该方法，但是这个方案存在一个比较明显的时间延迟，系统中平均响应时间达到 2-5 秒。

阿里团队所提出的分布式数据库架构(图 2)系统由客户端，根服务器，更新服务器，基线数据服务器和整合应用服务器五部分组成。

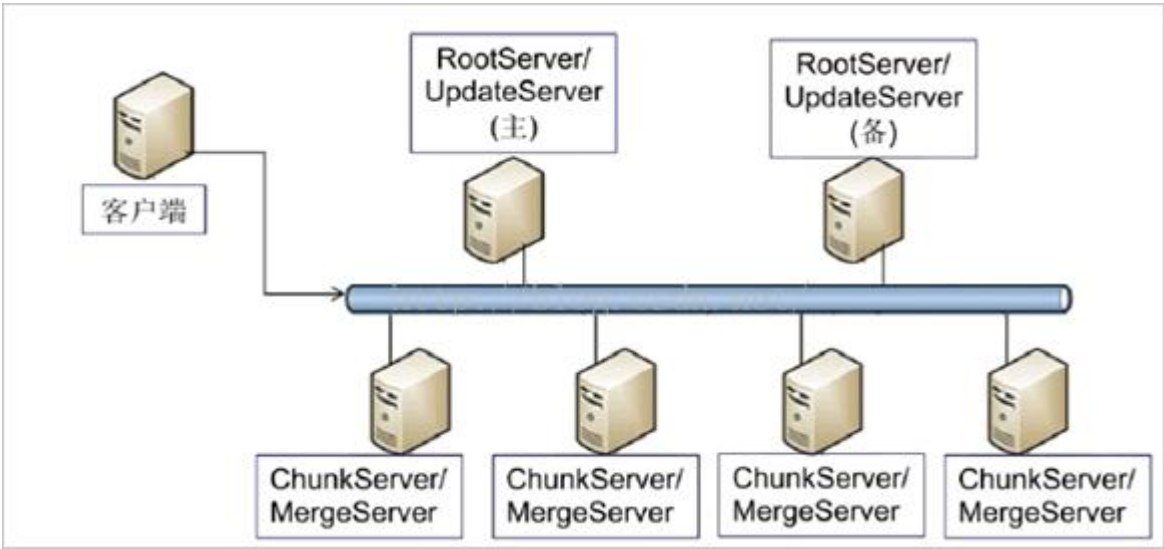


图 2 阿里团队分布式数据库架构图

根服务器用来管理集群中的所有物理主机，负责子表数据分布以及副本管

理，一般为一主一备，实现主备之间数据强同步；更新服务器用来保存系统的增量更新的数据，一般为一主一备，主备之间可以配置不同的同步模式，部署时常常和根服务器的进程共享物理服务器。基线数据服务器保存的是各个分片的数据集，存储系统中的基准数据，同时，基准数据往往是两份或多份；应用服务器则是接收并解析用户的 SQL 请求，并分发该请求。阿里团队基于上述的分布式数据库架构，阿里团队结合两阶段提交以及利用每台主机的事务控制功能，通过类二阶段提交协议，实现了双机一致性方案，但是实现多机分布式事务，这样的方式会使整个系统的复杂度大大提高，系统稳定性和并行处理能力也将受到考验。

根据阿里自身的业务需要，阿里团队采用单台更新服务器来记录最近一段时间的修改增量，而保存以往数据的基准服务器中的数据保持不变。在未同步到基准服务器时，每次请求将基准数据服务器和更新服务器中的增量数据一同返回给客户端，从而使写事务集中在单台更新服务器之中，避免复杂的分布式事务，高效的完成了跨表事务。同时，更新服务器上的增量数据会定期分发到基准数据服务器，实现良好的拓展性。但是上述做法只是在当前条件下用一个比较迂回的方式实现了跨表事务，增量数据和基准数据一同返回给客户端，会加大客户端数据处理的压力，同时这种方法仅仅是处于阿里自身业务出发的解决方式，是否能够满足不同业务需求，功能要求还需要结合实际情况进行进一步测试。

(3) 高可用和强同步

在实际生产环境之中，都需要一个高可用的方案来保证系统不间断运行，数据库作为系统数据的存储和服务核心能力，其可用性的要求要高于计算服务资源。目前高可用方案通常是多个数据库服务协同工作，在正常情况之下，系统会自动选择其中一个数据库作为主机。系统会向主机请求，所有数据的写操作，都会在主机，主服务器上操作，当一个主机发生故障时，余下的立刻顶替上去工作，这样子客户端请求，数据库发生故障时，可以做到不中断或者中断短时间。另外也可以是让多个数据库同时提供服务，用户通过路由分发，自动选择跳转路径短或者访问压力小的数据库来访问，当其中一个数据库故障时，立刻变更到另外的数据库进行访问即可。

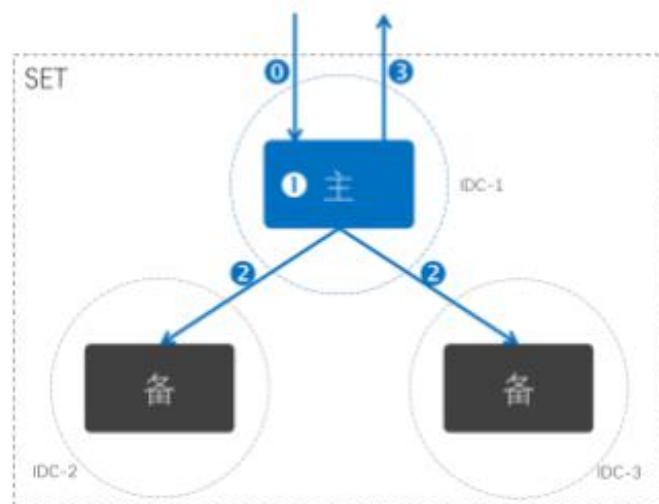


图 3 主备服务器模型

由于数据库记录了数据，要想要在多个数据库中进行切换，数据必须是同步的。数据库的同步是数据库高可用性的基础。目前数据同步的方式有三种：

一是异步复制，客户端发起更新数据的请求，主机完成相应操作之后，立即响应并通知备份数据库进行异步复制数据。但是这种方式当备份数据库发生故障时不会影响主机的工作，反之，则可能会导致数据不一致的情况出现；

二是强同步复制，即是客户端发起更新请求之后，主机完成相应操作，通知备份数据库开始同步数据，当接收到备份数据库的成功备份消息之后，主机再响应客户端。这种强同步模式下，主备数据库数据是同步进行的，备份数据库故障会影响主机操作，主机的故障却不会导致数据不一致，但是如果主备份数据库联通的网络出现问题，或者备份数据库数据出现问题，主机也会受到影响。同时，如果仅仅是一主一备的模式，也是做不到高可用的。

三是半同步复制，这是 Google 提出的一种折中的同步方式，正常情况下实行的是强同步复制，在备份机不可用的情况下，则会退化为异步复制，当异常恢复时，又会恢复为强同步模式。

5 分布式数据库未来发展趋势

随着业务不断云化，数据库也会随着业务不断云化，未来的业务可能都会“漫步在云端”，不管是私有还是公有，运维团队将接触到的可能不再是真实的物理机，而是一个个隔离的数据库节点。一个大的数据库承载着一切的业务，数据在底层打通，上层通过权限控制，节点控制等对数据切片，实现隔离，但是数据的

聚合和拓展将会变得简单，未来的业务层可能不再需要关心物理主机的容量和拓扑，只需要将其视作是一个近乎无穷大的数据库云平台，不用担心单机存储的容量和性能问题，负载均衡的问题。

未来的分布式数据库系统上，主从库同步，日志同步这样的备份方式将会被更强大的分布式一致性算法如 Raft，等所代替。人工的数据库运维在管理大规模的数据库集群时将不复存在，所有的容灾恢复，故障恢复，都将会是高度自动化。

用户在存储和读取数据时，将会变得更加简单，未来的分布式数据库将会实现更加高效便捷的存取方式，也许使用同一套标准语法和规则对数据进行读写和分析，将会给用户带来更好的体验。简言之，未来的分布式数据库将会是更加安全，敏捷，智能化。

参考文献

1. 数据库原理及应用
2. 腾讯云分布式数据库解决方案
3. 淘宝分布式数据库如何实现高可用