加餐1 概念解析：云原生、HTAP、图与内存数据库

我们课程的主要内容已经介绍完了，经过 24 讲的学习，相信你已经掌握了现代分布式数据库的核心内容。前面我所说的分布式数据库主要针对的是 NewSQL 数据库。而这一篇加餐，我将向你介绍一些与分布式数据库相关的名词和其背后的含义。学习完本讲后，当你再看见一款数据库时，就能知道其背后的所代表的意义了。

我首先介绍与 NewSQL 数据库直接相关的云原生数据库；其次会介绍目前非常火热的 HTAP 融合数据库概念；数据库的模式除了关系型之外，还有多种其他类型，我将以图数据库为例，带你领略非关系型数据库面对典型问题所具备的优势；最后为你介绍内存型数据库。

云原生数据库

云原生数据库从名字看是从云原生概念发展而来的。而云原生一般包含两个概念：一个是应用以服务化或云化的方式提供给用户；另一个就是应用可以依托云原生架构在本地部署与 Cloud 之间自由切换。

第一种概念是目前广泛介绍的。此类云原生数据库是在云基础架构之上进行构建的，数据库组件在云基础设施之上构建、部署和分发。这种云原生属性是它相比于其他类型数据库最大的特点。作为一种云平台，云原生数据库以 PaaS（平台即服务，Platform-as-a-Service）的形式进行分发，也经常被称作 DBaaS（数据库即服务，DataBase-as-a-Service）。用户可以将该平台用于多种目的，例如存储、管理和提取数据。

使用此类云原生数据库一般会获得这样几个好处。

快速恢复

也就是数据库在无须事先通知的情况下，即时处理崩溃或启动进程的能力。尽管现在有先进的技术，但是像磁盘故障、网络隔离故障，以及虚拟机异常等，仍然不可避免。对于传统数据库，这些故障非常棘手。因为用单个机器运行整个数据库，即便一个很小的问题都可能影响所有功能。

而云原生数据库的设计具有显著的快速恢复能力，也就是说可以立即重启或重新调度数据库工作负载。实际上，易处置性已从单个虚拟机扩展到了整个数据中心。随着我们的环境持续朝着更加稳定的方向发展，云原生数据库将发展到对此类故障无感知的状态。

安全性

DBaaS 运行在高度监控和安全的环境里，受到反恶意软件、反病毒软件和防火墙的保护。除了全天候监控和定期的软件升级以外，云环境还提供了额外的安全性。相反，传统数据库容易遭受数据丢失和被不受限制的访问。基于服务提供商通过即时快照副本提供的数据能力，用户可以达成很小的 RPO 和 RTO。

弹性扩展性

能够在运行时进行按需扩展的能力是任何企业成长的先决条件。因为这种能力让企业可以专注于追求商业目标，而不用担心存储空间大小的限制。传统数据库将所有文件和资源都存储在同一主机中，而云原生数据库则不同，它不仅允许你以不同的方式存储，而且不受各种存储故障的影响。

节省成本

建立一个数据中心是一项独立而完备的工程，需要大量的硬件投资，还需要能可靠管理和维护数据中心的训练有素的运维人员。此外，持续地运维会给你的财务带来相当大的压力。而使用云原生的 DBaaS 平台，你可以用较低的前期成本，获得一个可扩展的数据库，这可以让你腾出双手，实现更优化的资源分配。

以上描述的云原生数据库一般都是采用全新架构的 NewSQL 数据库。最典型的代表就是亚马逊的 Aurora。它使用了日志存储实现了 InnoDB 的功能，从而实现了分布式条件下的 MySQL。目前各个主要云厂商的 RDS 数据库都有这方面的优势，如阿里云 PolarX，等等。

而第二种云原生数据库理论上可以部署在企业内部（私有云）和云服务商（公有云），而许多企业尝试使用混合模式来进行部署。面向这种场景的云原生数据库并不会自己维护基础设施，而是使自己的产品能运行在多种云平台上，ClearDB 就是这类数

据库典型代表。这种跨云部署提高了数据库整体的稳定性，并可以改善服务全球应用的数据响应能力。

当然，云原生数据库并不仅限于关系型，目前我们发现 Redis、MongoDB 和 Elasticsearch 等数据库都在各个主要云厂商提供的服务中占有重要的地位。可以说广义上的云原生数据库含义是非常宽泛的。

以上带你了解了云原生数据库在交付侧带来的创新，下面我来介绍 HTAP 融合数据库在数据模型上有哪些新意。

HTAP

介绍 HTAP 之前，让我们回顾一下 OLTP 和 OLAP 的概念。我之前已经介绍过，OLTP 是面向交易的处理过程，单笔交易的数据量很小，但是要在很短的时间内给出结果；而 OLAP 场景通常是基于大数据集的运算。它们两个在大数据时代就已经分裂为两条发展路线了。

但是 OLAP 的数据往往来自 OLTP 系统，两者一般通过 ETL 进行衔接。为了提升 OLAP 的性能，我们需要在 ETL 过程中进行大量的预计算，包括数据结构的调整和业务逻辑处理。这样的好处是可以控制 OLAP 的访问延迟，提升用户体验。但是，因为要避免抽取数据对 OLTP 系统造成影响，所以必须在系统访问的低峰期才能启动 ETL 过程，例如对于电商系统，一般都是午夜进行。

这样一来， OLAP 与 OLTP 的数据延迟通常就在一天左右，大家习惯把这种时效性表述为 N+1。其中，N 就是指 OLTP 系统产生数据的日期，N+1 是 OLAP 中数据可用的日期，两者间隔为 1 天。你可能已经发现了，这个体系的主要问题就是 OLAP 系统的数据时效性，24 个小时的延迟太长了。是的，进入大数据时代后，商业决策更加注重数据的支撑，而且数据分析也不断向一线操作渗透，这都要求 OLAP 系统更快速地反映业务的变化。

此时就出现了将两种融合的 HTAP。HTAP (Hybrid Transaction/Analytical Processing) 就是混合事务分析处理，它最早出现在 2014 年 Gartner 的一份报告中。Gartner 用 HTAP 来描述一种新型数据库，它打破了 OLTP 和 OLAP 之间的隔阂，在一个数据库系统中同时支持事务型数据库场景和分析型数据库场景。这个构想非常美妙，HTAP 可以省去烦琐的 ETL 操作，避免批量处理造成的滞后，更快地对最新数据进行分析。

由于数据产生的源头在 OLTP 系统，所以 HTAP 概念很快成为 OLTP 数据库，尤其是 NewSQL 风格的分布式数据库和云原生数据库。通过实现 HTAP，它们试图向 OLAP 领域进军。这里面比较典型代表是 TiDB，TiDB 4.0 加上 TiFlash 这个架构能够符合 HTAP 的架构模式。

TiDB 是计算和存储分离的架构，底层的存储是多副本机制，可以把其中一些副本转换成列式存储的副本。OLAP 的请求可以直接打到列式的副本上，也就是 TiFlash 的副本来提供高性能列式的分析服务，做到了同一份数据既可以做实时的交易又做实时的分析，这是 TiDB 在架构层面的巨大创新和突破。

以上就是 HTAP 数据库的演化逻辑和典型代表。下面我们继续拓展知识的边界，看看多种模式的分布式数据库。

内存数据库

传统的数据库通常是采用基于磁盘的设计，原因在于早期数据库管理系统设计时受到了硬件资源如单 CPU、单核、可用内存小等条件的限制，把整个数据库放到内存里是不现实的，只能放在磁盘上。可以说内存数据库是在存储引擎层面上进行全新架构的 NewSQL 数据库。

伴随着技术的发展，内存已经越来越便宜，容量也越来越大。单台计算机的内存可以配置到几百 GB 甚至 TB 级别。对于一个数据库应用来说，这样的内存配置已经足够将所有的业务数据加载到内存中进行使用。通常来讲，结构化数据的规模并不会特别大，例如一个银行 10 年到 20 年的交易数据加在一起可能只有几十 TB。这样规模的结构化数据如果放在基于磁盘的 DBMS 中，在面对大规模 SQL 查询和交易处理时，受限于磁盘的 I/O 性能，很多时候数据库系统会成为整个应用系统的性能瓶颈。

如果我们为数据库服务器配置足够大的内存，是否仍然可以采用原来的架构，通过把所有的结构化数据加载到内存缓冲区中，就可以解决数据库系统的性能问题呢？这种方式虽然能够在一定程度上提高数据库系统的性能，但在日志机制和更新数据落盘等方面仍然受限于磁盘的读写速度，远没有发挥出大内存系统的优势。内存数据库管理系统和传统基于磁盘的数据库管理系统在架构设计和内存使用方式上还是有着明显的区别。

一个典型的内存数据库需要从下面几个方面进行优化。

1. 去掉写缓冲：传统的缓冲区机制在内存数据库中并不适用，锁和数据不需要再分两个地方存储，但仍然需要并发控制，需要采用与传统基于锁的悲观并发控制不同的并发控制策略。
2. 尽量减少运行时的开销：磁盘 I/O 不再是瓶颈，新的瓶颈在于计算性能和功能调用等方面，需要提高运行时性能。
3. 可扩展的高性能索引构建：虽然内存数据库不从磁盘读数据，但日志依然要写进磁盘，需要考虑日志写速度跟不上的问题。可以减少写日志的内容，例如把 undo 信息去掉，只写 redo 信息；只写数据但不写索引更新。如果数据库系统崩溃，从

磁盘上加载数据后，可以采用并发的方式重新建立索引。只要基础表在，索引就可以重建，在内存中重建索引的速度也比较快。

综上可以看到，构建内存数据库并不是简单地将磁盘去掉，而是需要从多个方面去优化，才能发挥出内存数据库的优势。而图数据库则是通过修改数据模型来实现高性能的，下面让我们看看其具体的功能。

图数据库

图数据库（graph database）是一个使用图结构进行语义查询的数据库，它使用节点、边和属性来表示和存储数据。该系统的关键概念是图，它直接将存储中的数据项，与数据节点和节点间表示关系的边的集合相关联。这些关系允许直接将存储区中的数据连接在一起，并且在许多情况下，可以通过一个操作进行检索。图数据库将数据之间的关系作为优先级。查询图数据库中的关系很快，因为它们永久存储在数据库本身中。可以使用图数据库直观地显示关系，使其对于高度互联的数据非常有用。

如果内存数据库是在底层存储上进行的创新，那么图数据库就是在数据模型上进行了改造。构造图数据库时，底层存储既可以使用 LSM 树等 KV 存储，也可以使用上一讲介绍的内存存储，当然一些图数据库也可以使用自创的存储结构。

由于图数据所依赖的算法本质上没有考虑分布式场景，故在分布式系统处理层面形成了两个流派。

以节点为中心

这是传统分布式理论作用在图数据库中的一种形式。此类数据库以节点为中心，使相邻的节点就近交换数据。图算法采用了比较直接的模式，其好处是并发程度较好，但是效率很低。适合于批量并行执行简单的图计算。典型代表是就是 Apache Spark。故此种数据库又被称为图计算引擎。

以算法为中心

这是针对图计算算法特别设计的数据库。其底层数据格式满足了算法的特性，从而可以使用较低的资源处理较多的图数据。但是此类数据库是很难进行大范围扩展的。其典型代表有 Neo4j。

当然我们一般可以同时使用上面两种处理模式。使用第一种来进行大规模数据的预处理和数据集成工作；使用第二种模式来进行图算法实现和与图应用的对接工作。

总之，图数据库领域目前方兴未艾，其在社交网络、反欺诈和人工智能等领域都有非常大的潜力。特别是针对最近的新冠疫情，有很多团队利用图数据库分析流行病学，从而使人们看到了该领域数据库的优势。

图数据库与文档型数据库、时间序列数据库等都是面向特殊数据模型的数据库。此类数据库之所以能越来越火热，除了它们所在领域被重视以外，最重要的底层原因还是存储引擎和分布式系统理论、工具的日趋成熟。可以预见，今后一些热门领域会相继产出具有领域特色的数据库。

总结

本讲作为加餐的知识补充，为你介绍了多种分布式数据库。其中云原生与 HTAP 都是 NewSQL 数据库的发展分支。云原生是从交付领域入手，提供给用户一个开箱即用的数据库。而 HTAP 拓展了 NewSQL 的边界，引入了 OLAP，从而抢占了部分传统大数据和数据分析领域的市场。

而内存数据库是针对数据库底层的创新，除了内存数据库外，2021 年随着 S3 存储带宽的增加和其单位存储价格持续走低，越来越多的数据库开始支持对象存储。今后，随着越来越多的创新硬件的加入，我们还可能看到更多软硬结合方案数据库的涌现。

最后我介绍了图数据库，它作为特殊领域数据库在领域内取得成就，是通用关系型数据无法企及的。而随着存储引擎和分布式理论的发展，此类数据库将会越来越多。

如此多的分布式数据库真是让人眼花缭乱，那么我们该如何去选择呢？下一节加餐我会结合几个典型领域，来给你一些指引。

上一页 下一页