加餐2 | 数据库选型：我们该用什么分布式数据库?

经过 24 讲的基础知识学习以及上一讲加餐，我已经向你介绍了分布式数据库方方面面的知识。这些知识，我觉得大概会在三个方面帮到你，分别是数据库研发、架构思维提升和产品选型。特别是 NewSQL 类数据库相关的知识，对于你认识面向交易的场景很有帮助。

我先后在电信与电商行业有十几年的积累，我想，你也非常希望知道在这些主流场景中分布式数据库目前的应用状况。这一讲我就要为你介绍银行、电信和电商领域内分布式数据库的使用状况。

首先我要从我的老本行电商行业开始。

电商：从中间件到 NewSQL

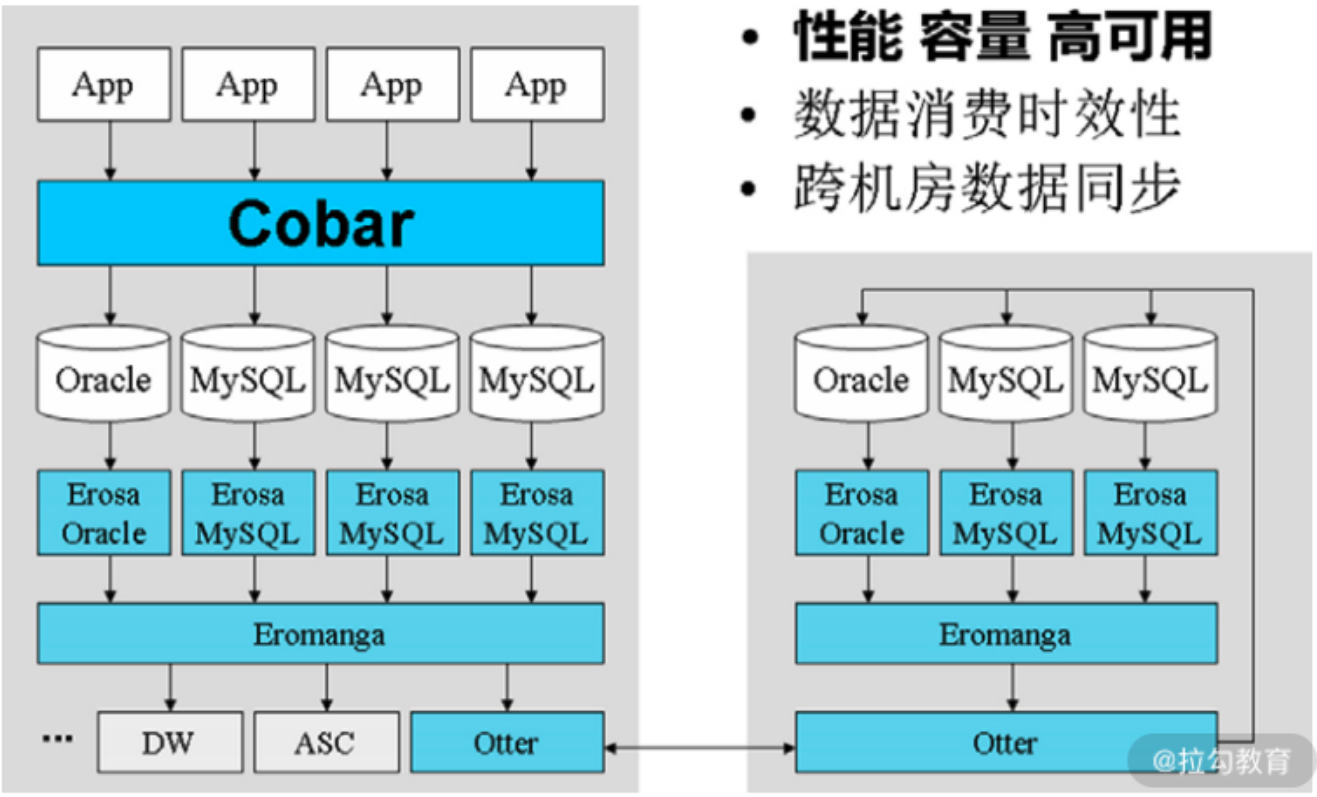
分布式数据库，特别是分布式中间这一概念就是从电商，尤其是阿里巴巴集团催生出来的。阿里集团也是最早涉及该领域的企业。这里我们以其 B2B 平台部产生的 Cobar 为例介绍。

2008 年，阿里巴巴 B2B 成立了平台技术部，为各个业务部门的产品提供底层的基础平台。这些平台涵盖 Web 框架、消息通信、分布式服务、数据库中间件等多个领域的产品。它们有的源于各个产品线在长期开发过程中沉淀出来的公共框架和系统，有的源于对现有产品和运维过程中新需求的发现。

数据库相关的平台就是其中之一，主要解决以下三方面的问题：

1. 为海量前台数据提供高性能、大容量、高可用性的访问；
2. 为数据变更的消费提供准实时的保障；
3. 高效的异地数据同步。

如下面的架构图所示，应用层通过 Cobar 访问数据库。



其对数据库的访问分为读操作（select）和写操作（update、insert和delete）。写操作会在数据库上产生变更记录，MySQL 的变更记录叫 binlog，Oracle 的变更记录叫 redolog。Erosa 产品解析这些变更记录，并以统一的格式缓存至 Eromanga 中，后者负责管理变更数据的生产者、Erosa 和消费者之间的关系，负责跨机房数据库同步的 Otter 是这些变更数据的消费者之一。

Cobar 可谓 OLTP 分布式数据库解决方案的先驱，至今其中的思想还可以从现在的中间件，甚至 NewSQL 数据库中看到。但在阿里集团服役三年后，由于人员变动而逐步停止维护。这个时候 MyCAT 开源社区接过了该项目的衣钵，在其上增加了诸多功能并进行 bug 修改，最终使其在多个行业中占用自己的位置。

但是就像我曾经介绍的那样，中间件产品并不是真正的分布式数据库，它有自己的局限。比如 SQL 支持、查询性能、分布式事务、运维能力，等等，都有不可逾越的天花板。而有一些中间件产品幸运地得以继续进阶，最终演化为 NewSQL，甚至是云原生产品。阿里云的 PolarDB 就是这种类型的代表，它的前身是阿里云的分库分表中间件产品 DRDS，而 DRDS 来源于淘宝系的 TDDL 中间件。

PolarDB 相比于传统的中间件差别是采用了共享存储架构。率先采用这种架构的恰好也是一家电商到云计算的巨头：亚马逊。而这个数据库产品就是 Aurora。

Aurora 采用了这样一种架构。它将分片的分界线下移到事务及索引系统的下层。这个时候由于存储引擎保留了完整的事务系统，已经不是无状态的，通常会保留单独的节点来处理服务。这样存储引擎主要保留了计算相关逻辑，而底层存储负责了存储相关的像 redo、刷脏以及故障恢复。因此这种结构也就是我们常说的计算存储分离架构，也被称为共享盘架构。

PolarDB 在 Aurora 的基础上采用了另外一条路径。RDMA 的出现和普及，大大加快了网络间的网络传输速率，PolarDB 认为未来网络的速度会接近总线速度，也就是瓶颈不再是网络，而是软件栈。因此 PolarDB 采用新硬件结合 Bypass Kernel 的方式来实现高效的共享盘实现，进而支撑高效的数据库服务。由于 PolarDB 的分片层次更低，从而做到更好的生态兼容，这也就是为什么 PolarDB 能够很快做到社区版本的全覆盖。副本件 PoalrDB 采用了 ParalleRaft 来允许一定范围内的乱序确认、乱序 Commit 以及乱序 Apply。但是 PolarDB 由于修改了 MySQL 的源码和数据格式，故不能与 MySQL 混合部署，它更适合作为云原生被部署在云端提供 DBaaS 服务。

以 Spanner 为代表的“无共享”类型的 NewSQL 由于有较高的分片层次，可以获得接近传统分库分表的扩展性，因此容易在 TPCC 这样的场景下取得好成绩，但其需要做业务改造也是一个大的限制。以 Aurora 及 PolarDB 为代表的“共享存储”的 NewSQL 则更倾向于良好的生态兼容，几乎为零的业务改造，来交换了一定程度的可扩展性。

可以说从 Cobar 到现在的 PolarDB，我们看到了分布式数据库在电商与云计算领域的成长。其他典型代表如京东数科的 ShardingShpere 也有类似的发展脉络。可以说电商领域，乃至互联网行业，都进入了以自主知识产权研发的模式，对技术的掌控力较高，从而打造了良好的生态环境。

以上介绍完了电商这个比较技术激进的行业，下面再来看一些传统行业的表现。

电信：去 IOE

电信作为重要的 IT 应用行业，过去 20 年一直引领着技术发展的潮流。但电信却又是关系到国计民生的重要行业，故技术路线相比于电商、互联网来说更为保守。其中 Oracle 数据库常年“霸占”着该领域，而分布式计算场景被众多子系统所承担着。以中国联通为例，在 2010 年前后，联通集团的各个省系统就有数百个之多，它们之间的数据是通过 ETL 工具进行同步，也就是从应用层保证了数据一致性。故从当时的情况看，其并没有对分布式数据库有很强的需求。

但在 2012 年前后，也就是距今大概 10 年前。中国联通开始尝试引入阿里集团的技术，其中上文提到的 Cobar 就在其中。为什么三大运营商中联通首先发力呢？原因是联通当年刚完成对老网通的合并，急需将移动网业务与固网业务进行整合。其次是联通集团总部倾向于打造全国集中系统，而这就需要阿里集团提供帮助。

根据当年参与该项目的人回忆，现场近千人一起参与，众多厂商系统工作，最终在距离承诺上线日期没几天的时候才完成了主要功能的验证。按现在的技术评估标准，当年这个项目并不是十分成功。但是这个项目把阿里的“去 IOE”理念深深地写入了电信领域内所有参与者的基因里面，从运营商到供应商，所有人都觉得分布式数据库的解决方案是未来的趋势。

而后 Cobar 衍生的 MyCAT 在联通与电信多个产品线开始落地，而各个供应商也开始模仿 Cobar 制作了自己的中间件产品。可以说，正是上面描述的背景，使数据库中间件模式逐步在电信领域内被广泛接受，其底色就是去“IOE”。

2016 年之后，电信行业没有放弃演化的步伐，各个主要供应商开始尝试去构建 NewSQL 数据库，其中特别是以 PGXC 架构为核心的 NewSQL 数据库引人瞩目。PGXC（PostgreSQL-XC）的本意是指一种以 PostgreSQL 为内核的开源分布式数据库。因为 PostgreSQL 的影响力和开放的软件版权协议（类似 BSD），很多厂商在 PGXC 上二次开发，推出自己的产品。不过，这些改动都没有变更主体架构风格，所以我把这类产品统称为 PGXC 风格，如亚信的 AntDB、人大金仓的 KingbaseDB 都是这类数据库的典型代表。此类数据库开始在行业内部快速落地。

近些年，电信行业内部也逐步接触了最具创新性的 NewSQL，但是此类数据库选择范围很小。电信运营商更倾向于与国内厂商合作，如 TiDB 和 OceanBase 已经有在三大运营商和铁搭公司上线的案例。不过，我们可以发现，目前新一代 NewSQL 接管的系统不是属于创新领域，就是属于边缘业务，还未触及核心系统。但是我们可以认为，未来将会有更多的场景使用到创新性的 NewSQL 数据库。

最后再来说说银行系统。

银行：稳中前进

银行与电信是非常类似的行业，但是银行的策略会更保守。银行并没有在内部推动轰轰烈烈的去“IOE”运动，故其在 OLTP 领域一直使用传统数据库，如 Oracle 和 DB2。

一直到近 5 年开始，银行的 IT 架构才发生重大调整。比如行业标杆的工商银行的架构改造在 2018 年大规模落地，而调研和试点工作则在更早的 2016～2017 年。这个时点上，商用 NewSQL 数据库刚刚推出不久，而金融场景的种种严苛要求，注定了银行不会做第一个吃螃蟹的人，那么这种可能性就被排除了。同样，另一种 PGXC 风格的分布式数据库也正待破茧而出，反而是它的前身，“分布式中间件 + 开源单体数据库”的组合更加普及。

后来的结果是工行选择了爱可生开源的 DBLE + MySQL 的组合，选择 MySQL 是因为它的普及程度足够高；而选择 DBLE 则因为它是在 MyCat 的基础上研发，号称“增强版 MyCat”，由于 MyCat 已经有较多的应用案例，所以这一点给 DBLE 带来不少加分。这一模式在工行内部推行得很好，最终使 MySQL 集群规模达到上千节点。虽然表面看起来还是非常保守，因为同期的电信行业已经开始推进 PGXC 架构落地了。但是工行，乃至真个银行行业，也走上了“去 IOE”的道路，可以想到，整个行业也是朝着 NewSQL 数据库的方向前进的。而且目前工商银行已经宣布与 OceanBase 进行合作，这预示着行业中 NewSQL 化的浪潮即将来临。

相对于 OLTP 技术应用上的平淡，工行在 OLAP 方面的技术创新则令人瞩目。基本是在同期，工行联合华为成功研发了 GaussDB 200，并在生产环境中投入使用。这款数据库对标了 Teradata 和 Greenplum 等国外 OLAP 数据库。在工行案例的加持下，目前不少银行计划或者正在使用这款产品替换 Teradata 数据库。

总结

这篇加餐我为你总结了几个重点行业使用分布式数据库，特别是 OLTP 类数据库的情况。我们可以从中看到一些共同点：

1. 发展脉络都是单机数据库、中间件到 NewSQL，甚至电商领域开始做云计算；
2. 各个行业依据自己的特点进行发展，虽然基本都经过了这些阶段，但是它们之间是有技术滞后性的；
3. 先发展的行业带动了其他行业，特别是电商领域的技术被其他行业引用，达到了协同发展的效果；
4. 国产数据库崛起，近年采购的全新架构的 NewSQL 数据库，我们都可以看到国产厂家的身影，一方面得力于国有电信、银行等企业的政策支持，另一方面国产数据库的进步也是大家有目共睹的。

这三个具有典型性的行业，为我们勾画来整个分布式数据库在中国的发展。希望你能从它们的发展轨迹中汲取养分，从而能使用分布式数据库为自己的工作、学习助力。

以上就是最后一节加餐的内容了，希望可以帮助到你，再见。