

观点 | 云原生时代来袭 下一代云数据库技术将走向何方？

转载

xstardust 于 2019-04-24 16:02:49 发布 61 收藏

文章标签：[数据迁移](#) [分布式架构](#) [数据同步](#)

全面云化的时代已经到来，面对一系列的新技术和挑战，数据库市场将面临怎样的变革？作为云服务提供商，如何帮助更多的企业级用户把握“云”潮，提供最高效、最具价值的数据库解决方案？

日前，在阿里云峰会·北京站的数据库专场上，阿里巴巴集团副总裁、达摩院首席数据库科学家、阿里云智能事业群数据库产品事业部总负责李飞飞针对下一代云原生数据库的技术与挑战进行了精彩分享。



数据库发展与技术演进之路

根据DB-Engine 2019年1月份的数据库市场趋势分析，关系型数据库依旧占据着最核心的市场份额。与此同时，数据库市场也在不断细分，图数据库、文档数据库以及NoSQL等数据库细分市场正在崛起。

另一大趋势则是：以Oracle、DB2和Microsoft SQL Server三大巨头为代表的传统商业数据库所占据的市场份额不断下降，而开源和第三方数据库市场持续增长。

数据库技术诞生至今，虽然已经经历40多年的发展历程，但仍处于蓬勃发展的时期。如今，各大云计算厂商也达成了共识：数据库是连接IaaS和云上智能化应用的重要组成部分，因此从数据的产生、存储以及消费等各个环节，云厂商都需要提升全链路的能力，进而满足用户连接IaaS和智能化应用的需求。

随着数据库技术的不断发展，不仅出现了OLTP系统来实现事务的处理，可以实现对交易数据的实时记录；还出现了OLAP系统，借助OLAP系统可实现对于海量数据的实时分析。除此之外，还需要各类的数据库服务和管控工具支持核心的OLTP和OLAP系统。在此基础之上，针对半结构化数据和非结构化数据，NoSQL数据库解决方案应运而生。

上世纪70年代末到80年代初，诞生了关系型数据库，产生了SQL查询语言以及OLTP交易处理系统。随着数据的爆发式增长以及复杂分析需求的出现，诞生了数据仓库，以及OLAP在线分析处理系统以及ETL等数据处理技术。

技术发展到今天，图、文档以及时空、时序等多元异构数据量持续增长，因此也对应地出现了关系型数据库之外的NoSQL和NewSQL数据库系统。

云原生时代 我们需要怎样的数据库？

传统数据库往往采用单节点架构，而到了如今的云原生时代，云原生数据库通常会采用共享存储的架构方式。阿里云POLARDB所采用的就是共享存储架构，通过高速网络构建共享存储，在此之上实现存储与计算的分离，进而可以快速弹性扩容出多个计算节点。

与此同时，POLARDB还可以根据用户的具体需求在存储和计算两个维度迅速地实现扩缩容。而对于用户而言，无需修改任何业务逻辑，就可以使用基于共享存储的POLARDB数据库，可以实现无侵入式迁移。

除了云原生的共享存储技术之外，面对高并发、海量数据访问的挑战，也需要借助分布式架构来解决，比如为了应对每年的双11大促，阿里巴巴自身就需要去探索分布式架构。

此外，针对数据的多模多态需求，阿里云也希望在用户侧提供不同查询接口，比如SQL。在存储侧，阿里云希望能够支持用户将数据存储到不同地方，并通过像SQL这样的统一接口实现对不同数据类型的统一查询访问。目前，阿里云所提供数据湖服务就是针对上述场景所演进来的云原生技术。

正如前面所提到的OLTP和OLAP系统，传统的解决方案希望将读写冲突隔离开，让OLTP负责事务处理，让OLAP负责海量数据的分析任务。而在云原生时代，阿里云会借助新硬件所带来的技术红利，尽可能降低数据迁移的成本，将事务处理和数据分析整合在同一个引擎中，通过一套系统为用户无缝解决两种问题。

阿里云目前已经服务了大量企业级客户，这些客户通过虚拟化、存储与计算分离来使用阿里云所提供的云资源池。因此，我们需要对于云上全部资源进行智能化监控和调配，做到快速响应，并且为用户提供最高质量的服务。而智能化的背后需要利用机器学习以及人工智能技术，从数据迁移、数据保护、弹性调度等各个维度实现自感知、自决策、自恢复以及自优化。

在云原生时代，另外一个重要技术就是软硬件一体化设计。新硬件的发展为数据库系统带来了许多可不断利用的技术红利，比如RDMA网络、SSD、NVM、GPU、IPG加速等。阿里云的POLARDB共享存储就利用了RDMA网络，因此可以做到像本地节点一样快速访问远程数据库节点。

对于很多云上的客户而言，可能也有金融级高可用的需求。利用高可用协议，阿里云数据库可以采用三副本架构，在本地可以实现数据库之间的无缝实时切换，在异地也可以满足不同用户对于灾备的需求，借助Binlog技术实现异地数据同步，实现金融级高可用。此外，云上用户对数据的安全性尤为看重，阿里云数据库服务从数据落盘开始，就提供了加密技术，确保全链路的数据安全。

阿里云数据库服务：全域布局自主可控 兼具创新与商用价值

阿里云数据库所提供的工具类产品包括数据备份、数据迁移、数据管理、混合云数据管理以及智能化诊断与优化系统等，可以帮助客户实现快速上云并且打造混合云解决方案。

我们提供的核心引擎产品中，既有自主可控的自研产品，也有第三方以及开源的产品。希望通过商业数据库以及开源产品为用户提供丰富的选择，同时也希望将云计算的技术红利整合到自研数据库产品中去，进一步做深做透，真正地帮助客户解决应用第三方或者开源数据库所无法解决的痛点和问题。

在OLTP方向，阿里云数据库所提供的核心产品是POLARDB及其分布式版本POLARDB X。除此之外，阿里云也提供主流的MySQL、PostgreSQL、SQLServer以及针对于Oracle兼容的PPAS等一系列服务。

针对OLAP系统，我们的核心产品是AnalyticDB、针对多源异构数据所提供的数据湖服务DLA，以及针对IoT场景的时序时空数据库TSDB。而在NoSQL领域，阿里云则提供了丰富的第三方数据库产品供客户选择，比如HBase、Redis、MongoDB以及阿里云自研的图数据库GDB等。

阿里云数据库的管控平台与全链路监控服务为用户提供了智能化的全链路检测与分析，保障阿里云数据库能够为用户提供最高级别的Service Level。

阿里云帮助客户打造了一条线上线下混合云数据存储的链路，从客户迁移上云开始，其可以选择阿里云DTS服务进行实时的数据上传和同步。数据上云之后，客户可选择POLARDB等云原生数据库产品进行存储，也可以借助DLA或者AnalyticDB等产品进行数据分析。

而针对特定场景进行数据分析可以选用文档数据库、图数据库或者时序时空数据库等解决方案，可以借助DTS系统实现线上线下的数据同步备份，并且还可以借助HDM来进行混合云数据库管理。此外，阿里云数据库服务还提供了数据库管理套件，可以支撑用户对数据库进行管理和开发，使得管理和开发流程更为高效。

这里主要为大家介绍两种云原生数据库产品，POLARDB和AnalyticDB。



2019 阿里云峰会·北京
2019 ALIBABA CLOUD SUMMIT

POLARDB v2.0发布

New: POLARDB v2.0

兼容PostgreSQL Oracle

为互联网海量数据和海量并发而生

- Cloud-Native OLTP数据库
支持Oracle和PostgreSQL 11语法
- 88核 710G
超大实例规模
- 分钟级弹性
最低成本应对业务高峰
- 100TB
满足企业大数据存储需求
- SQL
洞察能力
- 20倍
更强大的优化器
更好的并行查询能力

https://blog.csdn.net/xstardust

POLARDB

POLARDB利用RDMA网络实现了高效的共享分布式存储，借助于共享存储技术，多个计算节点中可以实现“一写多读”，并且可以根据客户的工作量需求，快速弹出多个只读节点来满足客户在高峰时刻对于计算的需求，同时也可以存储在存储节点上实现快速缩扩容。

针对客户的应用场景以及其业务峰值峰谷的波动情况，POLARDB可以做到按量按需使用和计费，进而大大地提升了客户的数据库使用效率，节省了所需要成本。总体而言，POLARDB就是一个超级MySQL，后续POLARDB还会陆续推出兼容PostgreSQL和Oracle的版本。

在一些场景下，用户需要面对高并发和海量数据访问的挑战，因此需要突破共享存储的容量上限。分布式架构的POLARDB X则利用Sharding Partition解决方案实现了存储容量的无限水平扩展。POLARDB X分布式版本也会在后续进行公测，欢迎大家试用。

AnalyticDB

对海量数据进行分析的时候，读写会产生一定的冲突。如果需要读取大量数据并进行分析将会极为复杂，因此推荐大家使用阿里云的实时交互式分析数据库系统AnalyticDB。

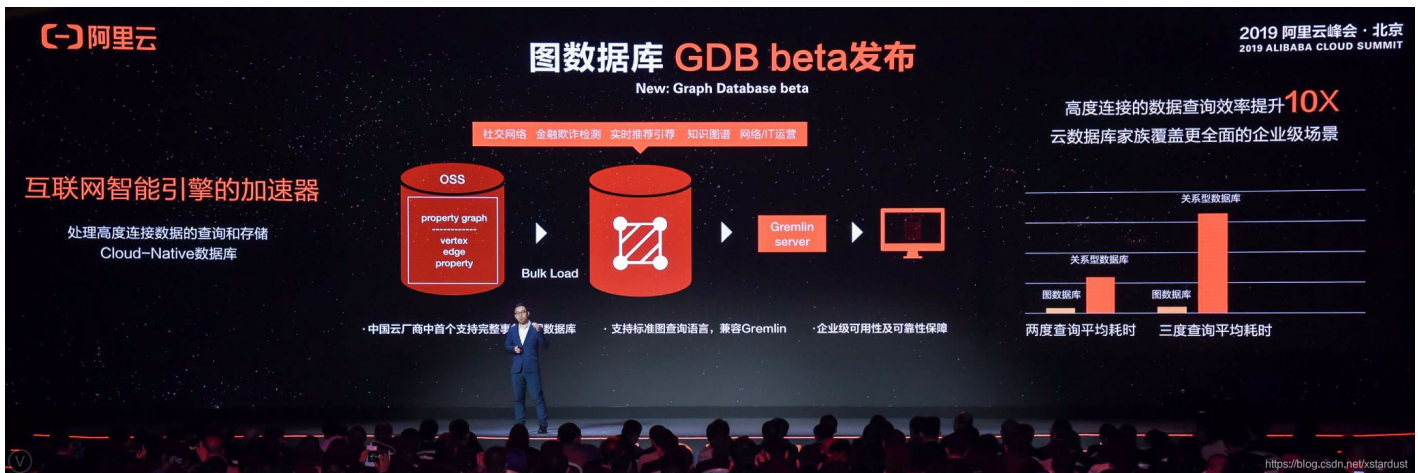
AnalyticDB最核心的特点是具有支持高吞吐写入的能力，并且具有针对于行列存储所研发的存储引擎，因此可以实现实时的交互式分析。在海量数据以及高并发场景下，AnalyticDB在响应时间方面的表现都非常优秀。AnalyticDB兼容了MySQL生态，因此可以将MySQL里面的数据直接导入AnalyticDB中，可以实现对上百亿级别的数据的毫秒级查询以及百万TPS级别的写入。

数据传输云服务DTS

除了核心的云原生数据库产品，我们还有多款数据库工具型产品，比如数据传输云服务DTS。DTS所解决的痛点是云下客户上云所需要进行的数据传输问题，以及其在上云之后，云上与云下数据库或者从TP到AP系统之间进行实时的数据同步问题。

利用DTS，用户可以实现快速、高效地增量数据同步，保证了实时数据一致。DTS还提供了数据订阅能力，可以通过不同的协议和接口来接入更多不同的数据源。

数据库家族迎来新成员：图数据库GDB



图数据库 GDB beta发布
New: Graph Database beta

2019 阿里云峰会·北京
2019 ALIBABA CLOUD SUMMIT

互联网智能引擎的加速器
处理高度连接数据的查询和存储
Cloud-Native数据库

OSS
property graph
vertex
edge
property

Bulk Load

Gremlin server

中国云厂商中首个支持完整图数据库 · 支持标准图查询语言，兼容Gremlin · 企业级可用性及可靠性保障

高度连接的数据查询效率提升10X
云数据库家族覆盖更全面的企业级场景

图数据库 关系型数据库
两度查询平均耗时 三度查询平均耗时

https://blog.csdn.net/xstardust

这里为大家介绍一款阿里云新的数据库产品——阿里云图数据库GDB，该产品目前正在阿里云官网进行公测。GDB是一种支持属性图模型，用于处理高度连接数据查询与存储的实时可靠的在线数据库，其利用了大量的云原生技术，比如存储与计算分离等。GDB支持标准的图查询语言，兼容Gremlin语法，这一点与市场主流的图数据库保持一致。

GDB另外一个最核心的特点就是支持实时更新、支持OLTP级别的数据一致性，能够帮助大家对海量的属性图进行分析存储时保证数据的一致性。GDB具有服务高可用、易于维护等云原生数据库产品的特性。比较典型的应用场景有社交网络、金融欺诈检测、实施推荐等，同时GDB还支持知识图谱等形态以及神经网络等。

以赋能客户为本：降本提效 无忧发展

阿里云数据库团队的目标是为客户提供企业级的云原生数据库服务，利用自身全域布局、自主可控的技术为企业客户提供快速数据上云、云上云下统一数据管理以及数据安全等服务。

举例而言，阿里云数据库服务目前在杭州等城市支撑了像城市大脑这样复杂的应用，其既需要存储结构化数据也需要存储非结构数据，并且对于OLTP、OLAP、工具类产品等都提出了巨大的挑战，而借助阿里云AnalyticDB、POLARDB、DTS等利用了云原生技术的产品可以无缝地支撑城市大脑这样的复杂应用场景。

以POLARDB为例，该产品于2018年8月份进行公测，2018年的年底进行商业化，到目前为止实现了在公有云平台上的快速增长。POLARDB呈现快速增长背后的核心原因就是阿里云真正地帮助客户解决自身的痛点问题，不是利用新技术来“造完锤子找钉子”，而是真正地“看到钉子再去造锤子”。

POLARDB最为核心的特点就是云原生、分钟级别的弹性存储和计算、高性价比、灵活弹性的使用计费方式、高并发能力，可快速扩容多个只读节点、大容量支持、通过共享分布式存储做到了类似单机数据库的体验，对用户的业务逻辑无侵入，并且高度兼容MySQL。

AnalyticDB则是一个实时交互式分析系统，无论是自制数据还是从大数据系统中获取的存储数据，都可以借助DTS工具将其迁移到AnalyticDB集群上，进行深入的商业分析、可视化以及交互式查询等。

AnalyticDB能够支持上百个表的连接查询，可为客户提供毫秒级的查询服务。如今，无数用户正在阿里云上使用POLARDB和AnalyticDB这样的云原生数据库，云原生数据库也正在真正地改变客户在应用中所遇到的痛点，为他们带来更多的业务价值。

总结而言，云原生时代出现了一系列的新技术和新挑战。面对这些挑战，需要对于数据库内核产品、管控平台以及数据库工具进行有机整合，才能够为客户提供最高效、最具有价值的解决方案。阿里云诚挚地邀请大家体验自身的产品和技术，希望能够与更多的客户一起合作，解决问题，也希望更多的开发者和生态合作伙伴能够基于阿里云的数据库服务和产品打造针对特定行业和领域的深度解决方案，使得云原生时代的数据库市场更加繁荣。

[原文链接](#)