

基于 NewSQL 的数据库发展现状及趋势分析

0121710880223 - 软件工程 1702 - 刘佳迎

日期: September 4, 2019

摘 要

未来 10 年是数据爆炸式增长的 10 年,当前正由 IT 时代进入 DT 时代。面对诸如大数据、人工智能、物联网等多元化应用场景,云数据库就是其背后的技术壁垒。因此本文基于市场发展趋势、技术领域创新等特点对数据库市场的三大阵营: *OldSQL*、*NewSQL* 和 *NoSQL* 进行对比分析,并从大数据发展的角度对数据库技术的发展做出了预测。

关键词: 数据库技术 NewSQL 数据库架构 列式存储技术

1 数据库市场发展趋势

1.1 大数据需求增长明显

数据库市场成形于上个世纪 80 年代。在过去 30 多年中,全球数据库市场基本由美国厂商(其数据库产品主要以事务处理为主)垄断,不仅中国数据库厂商很难找到突破口,就连德国和日本的厂商也举步维艰。随着云计算、大数据的兴起,中国数据库厂商迎来了发展的新机会。一方面,面对爆发式增长的海量数据,基于传统架构的信息系统已难以应对,同时传统商业智能系统和数据分析软件,面对以视频、图片、文字等非结构化数据为主的大数据时,也缺少有效地分析工具和方法。另一方面,大数据面临着有效存储、实时分析等挑战。在这种情况下,数据库行业迎来了新一轮的创新起点。而在大数据带来的创新“十字路口”,国外企业也处于战略布局的初始阶段,和国内企业同处创新起跑线上。

国家信息化建设快速推进、电子政务正处于深化应用阶段,对数据的开发利用将成为新的方向,同时出于信息安全的角度,国产数据库软件必将获得政府青睐。可以说,大数据给国产数据库厂商带来了“天时、地利、人和”三者同时齐备的三十年不遇的战略机遇期,如果国内企业能抓住这一机遇,在技术、商业模式上主动突破,形成自己的优势,在某些行业率先开辟出应用示范案例^[1],那么未来在市场上将会取得突破性的发展。

1.2 大数据成为厂商未来重要战略方向

大数据已经成为一种社会现象:它既是对既往 IT 技术发展至今的高度抽象和概括,同时抓住了 IT 技术服务于人、数据蕴藏价值的本质,因此大数据正日益成为社会各界关注的热点。目前,各研究机构以及 IBM、甲骨文、EMC 等信息技术企业都将大数据视为继云计算、物联网之

后信息技术领域的又一重点，并认为大数据的发展将带来可观的经济社会效益。大数据中蕴含的大商机是每个厂商都渴求抓住的机遇，如图 1 包括甲骨文、IBM 以及微软在内的主流数据库厂商都纷纷发布了明确的大数据战略。

企业名称	解决方案
甲骨文	甲骨文在2011年10月初召开的Oracle OpenWorld 2011大会上正式推出了Oracle大数据机（Oracle Big Data Appliance）。Oracle大数据机集成了硬件、存储和软件，包括Apache Hadoop软件的开源代码分发、新的甲骨文NoSQL数据库和用于统计分析的R语言开源代码分发。
IBM	2011年5月，IBM正式推出InfoSphere大数据分析平台。InfoSphere大数据分析平台包括BigInsights和Streams，还集成了数据仓库、数据库、数据集成、业务流程管理等组件。
微软	微软于2011年发布SQL Server R2 Parallel Data Warehouse（PDW，并行数据仓库），PDW使用了大规模并行处理来支持高扩展性，它可以帮助客户扩展部署数百TB级别数据的分析解决方案。
SAP	SAP Sybase IQ（最新版本15.4）是SAP公司面向大数据的高级分析平台，它经过专门设计，可为任何标准硬件和操作系统上的任务关键型商业智能、分析和数据仓库应用解决方案提供卓越的性能。
Teradata	Teradata推出一款集硬件、软件和服务于一体的全面产品组合——Teradata分析生态系统（Teradata Analytical Ecosystem）。Teradata Unity 将确保整个Teradata Analytical Ecosystem的同步和统一。
亚马逊	亚马逊在2009年推出了亚马逊弹性MapReduce（Amazon Elastic MapReduce）。
谷歌	Google推出Web服务BigQuery，用来在云端处理大数据。Google曾表示BigQuery引擎可以快速扫描高达70TB未经压缩处理的数据，并且可马上得到分析结果。
EMC	EMC公司推出用于支持大数据分析的下一代平台——EMC Greenplum统一分析平台（UAP）。
NetApp	Network Appliance, Inc.（NetApp）推出StorageGRID对象存储软件解决方案，设计用于管理PB级、全球分布的存储库。
惠普	惠普企业服务事业部宣布推出帮助客户更快部署惠普子公司Vertica的Vertica Analytics Platform。该产品让用户能够大规模实时分析物理、虚拟和云环境中的结构化、半结构化和非结构化数据，从而深入洞悉“大数据”。
沃尔玛	沃尔玛为进军移动互联网和挖掘大数据进行了一系列的收购：包括Kosmix（沃尔玛实验室前身）、Small Society、Set Direction、OneRiot、Social Calenda、Grabble等多家中小型创业公司，这些创业公司要么精于数据挖掘和各种算法，要么在移动社交领域有其专长。
英特尔	2012年7月英特尔公司对外发布了Hadoop商业发行版（Apache Hadoop Distribution）。

图 1 国外 IT 厂商的大数据解决方案概览

2 数据库技术发展现状

美国著名数据库科学家迈克尔·斯通布雷克 (Michael Stonebraker) 指出^[1]，行业技术的发展趋势^[4]是由一种架构支持所有应用转变为用多种架构支持多类应用。在大数据和云计算的背景下，这一理论导致了数据库市场的大裂变：数据库市场分化为三大阵营，包括 *OldSQL*（下称传

统数据库或 *OldSQL*)、*NewSQL* (下称新型数据库或 *NewSQL*) 和 *NoSQL* (下称非关系型数据库或 *NoSQL*)。

2.1 大数据引发数据库架构变革

传统数据库的基本架构是 30 年前以事务处理为主要应用设计的,可以概括为“一种架构支持多类应用”。然而,在大数据时代,对于分析类应用的需求不断增加,这种变化导致一种架构难以完全满足大数据的需求,“多种架构支持多类应用”成为数据库行业应对大数据的基本思路。于是,数据库行业出现三个互为补充的三大阵营,*OldSQL*、*NewSQL* 和 *NoSQL*。

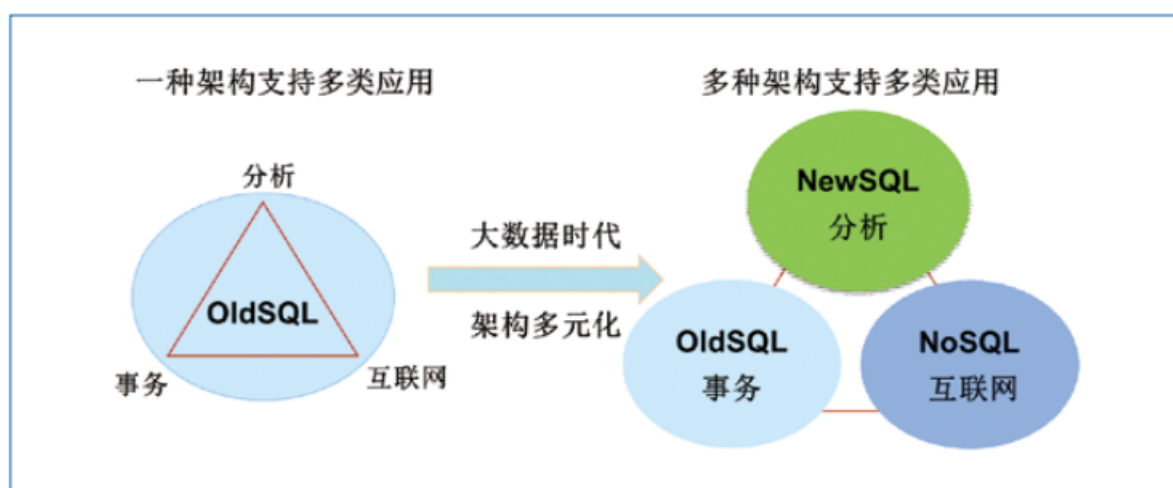


图 2 大数据引发的数据库架构变革

2.2 大数据推动数据库技术创新

从技术角度看,*OldSQL* 的典型特征是行存储、关系型和 *SMP* (对称多处理架构)^[2]。*OldSQL* 的代表产品包括 *TimesTen*、*Altibase*、*SolidDB* 和 *Exadata* 等。*OldSQL* 所代表的传统关系型数据库已经不能满足大数据对大容量、高性能和多数据类型的处理要求。为了更好地满足云计算和大数据的需求,*NewSQL* 和 *NoSQL* 脱颖而出,并且大有后来者居上的架势。*NoSQL* 的技术主要源于互联网公司^[3],如 *Google*、*Yahoo*、*Amazon*、*Facebook* 等。*NoSQL* 产品普遍采用了 *Key-Value*、*MapReduce*、*MPP* (大规模并行处理) 等核心技术。在互联网大数据应用中,*NoSQL* 占据了主导地位。

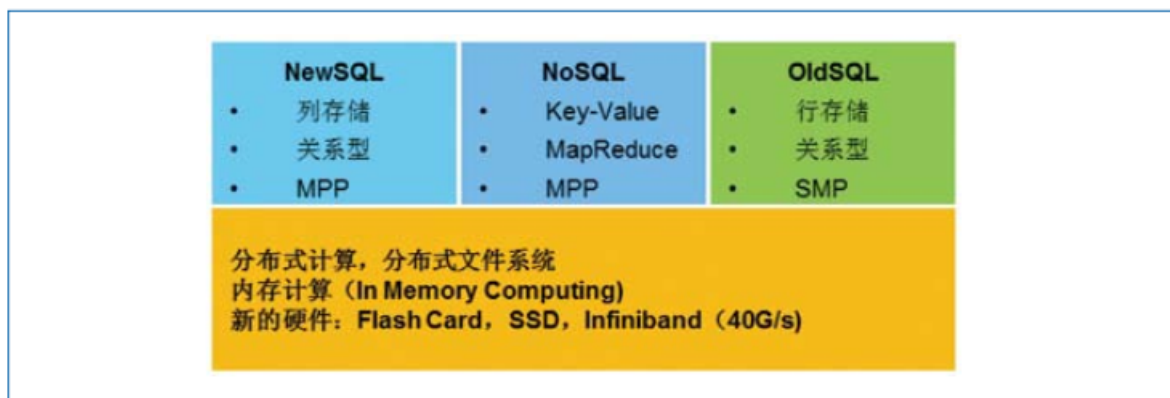


图 3 大数据带来的数据库技术创新

大数据激发了数据库行业技术创新的热情，主要的驱动力是对处理性能的强烈需求。为了提升性能，*NewSQL* 阵营普遍采用了列存储技术；*NoSQL* 阵营普遍采用了 *KV* 技术。三个阵营都不同程度地采用了分布式计算、分布式文件系统、内存计算技术，并积极地使用新的硬件技术，如大内存、*Flash*、*SSD* 和高速网络连接（万兆交换机和 *Infiniband*）等。

2.3 大数据激发数据库产品创新

大数据推动了数据库行业的产品创新，*NewSQL* 阵营在过去五年里形成了近十个商用的产品，谷歌发表论文介绍了 *F1/Spanner* 关系型数据库（未开源）。*NoSQL* 阵营的技术源于互联网公司谷歌、*Yahoo*、亚马逊、*Facebook* 等。*OldSQL* 阵营在基本架构不变的基础上引入内存计算和一体机技术以提升处理性能。

表 1 大数据带来的数据库产品创新示例

NewSQL	NoSQL	OldSQL
GBase 8a	Hadoop	TimesTen
Greenplum	HBase	Altibase
Vertica	Bigtable	SolidDB
AsterData	Cassandra	Exadata
Sybase IQ	Dynamo	Netezza
F1/Spanner	Dremel	Teradata

3 列式存储技术将成为大数据时代数据库技术的理想架构

传统的关系型数据库不易扩展，也不能实现并行处理，因此在处理海量数据时显得有些捉襟见肘。在大数据场景下，以“一种架构支持所有应用”的传统数据库势必会给用户带来使用上的不便，如强迫用户不断增加数据中心的面积，从而带来能源等方面的巨大消耗^[1]，即便在云计算技术的支持下，传统数据库也将迫使用户不断增加其对于云的需求，额外的增加开支。其次，通过行业大数据为数据库厂商带来了更大的商业机会的分析可以看到，在大数据场景下，*NoSQL* 则更多的被互联网厂商所采用。而给数据库厂商带来更大价值的则是新型数据库。而新型数据

库厂商大多都采用了列存技术，因此可以说，列式存储技术可以看做大数据时代数据库技术的理想架构。

参考文献

- [1] BENNETT R M, PICKERING M, SARGENT J. Transformations, transitions, or tall tales? a global review of the uptake and impact of nosql, blockchain, and big data analytics on the land administration sector[J]. Land Use Policy, 2019, 83: 435-448.
- [2] MEO R, PSAILA G, CERI S, et al. A new sql-like operator for mining association rules[C]//VLDB: volume 96. [S.l.: s.n.], 1996: 122-133.
- [3] HUTCHISON D, KEPNER J, GADEPALLY V, et al. From nosql accumulo to newsql graphulo: Design and utility of graph algorithms inside a bigtable database[C]//2016 IEEE High Performance Extreme Computing Conference (HPEC). [S.l.]: IEEE, 2016: 1-9.
- [4] HAN J, HAIHONG E, LE G, et al. Survey on nosql database[C]//2011 6th international conference on pervasive computing and applications. [S.l.]: IEEE, 2011: 363-366.
- [5] CATTELL R. Scalable sql and nosql data stores[J]. Acn Sigmod Record, 2011, 39(4):12-27.