



控制NoSQL数据库

数据驱动型企业必须采用新的数据建模思路

更多信息，请访问erwin.com



第1节

▶NoSQL对数据驱动型组织至关重要

世界正在以前所未有的速度发展，因而企业需要紧跟时代步伐，成为数据驱动型组织。很多公司都在努力实现这一目标，目的是提供流畅的按需数据访问，以便能够在上下文中为关键决策提供信息并产生影响，从而改善业务成果。

如果企业想要成为一个能够有效地实时响应客户、合作伙伴、供应商和其他各方并从这些工作中获利的数字企业，那么加快大数据（其中大部分是非结构化数据）检索和分析的速度至关重要。

能获得哪些收益？专家评估表明，与同行相比，使用数据驱动型决策的企业的**生产率和利润率高出5%**。能够有效利用分析方法的企业**每投资1美元就能获得13美元的回报**。

然而，要想成为数据驱动型组织，企业就必须不断

提高其利用NoSQL解决方案的能力，以便将非结构化数据用于入微的洞察并提供算法，从而采取最合适的行动。

事实上，如果企业能够以很快的速度利用和查询NoSQL数据库中的大量非结构化、结构化和半结构化数据，那么这些数据就能够成为支持现代云应用及其规模、速度和敏捷开发需求的重要资产。

这一点尤其适用于全球发展最快的NoSQL数据库平台MongoDB。

1

预计**2017**年的大数据投资额将超过**570**亿美元。

未来三年的投资复合年增长率约为**10%**。

▶资料来源：SNS研究

2

81%的公司理解数据对于提高效率和业务绩效的重要性。

但只有**3%**的公司认为其数据策略“成熟”。

▶资料来源：EY（与Nimbus Ninety合作）

3

到**2020**年，NoSQL的市场价值预计将达到**42**亿美元。

▶资料来源：Allied Market Research

4

90%的数据是非结构化数据。

▶资料来源：Gartner

►NoSQL的兴起改变了建模格局

当然，结构化数据和关系数据库在数据驱动型企业中仍然发挥着各自的作用。如果一家公司发现采用新颜色的产品销售业绩不佳，就可以将有关此结果的结构化数据与非结构化数据（例如社交媒体评论）相关联，从而发现该产品不受买家青睐的原因。

虽然SQL世界和NoSQL世界可以在当今的大数据生态系统中相互补充，但大多数企业需要集中力量为后一种格式构建专业知识和流程。

毕竟，他们数十年来一直从事SQL数据库的设计和管理，这些工作重视存储效率和引用完整性，而不是快速数据访问，而后者对构建能够向员工、客户

和其他各方提供实时价值的云应用程序非常重要。

当涉及到支持当今快速交付的迭代型实时应用程序时，针对查询优化的建模就成了新的口号（请参阅边栏《提高数据访问速度的新方法》）。

在MongoDB等NoSQL平台中，建模的面貌焕然一新。NoSQL平台本质上更为灵活，因此有利于提高速度、促进敏捷开发、改善水平伸缩性、简化设计和提高可用性，同时改善控制并降低成本。

与此同时，NoSQL强调开发人员对数据建模人员的管理，而现在我们应该更好地平衡二者的关系。

采用面向NoSQL数据建模的新思路的用例包括：

- 将现有关系数据库模型转换为使用本机非关系建模支持的NoSQL模型，作为面向云环境的刷新应用程序的一部分；
- 对现有NoSQL实施进行反向工程，这些实施的特点是在应用程序代码本身内部实施结构，以供审查和获取知识；以及
- 从头开始为新应用程序构建NoSQL模型。

提高数据访问速度的新方法

Query-Optimized Modeling™可以帮助数据专业人员和开发人员从传统的关系结构迁移到基于云的NoSQL集合结构，从而有效地构建能够支持新应用程序数据访问需求的模式。erwin正在申请专利的技术可用于从针对存储优化的模型迁移到针对查询优化的模型，以提供最轻松最快捷的数据获取路径，从而满足应用程序和开发人员的要求。

利用这种能力的人可以决定使用标准化方式进行转换，这样可以使模型尽可能接近关系数据模型，并专门用于对顶级集合的差异化查询和交叉引用。

或者，他们可以利用反标准化方法将数据复制到任何有需要的位置。

这种方法会减少集合之间的连接，通常更有利于提高访问速度。

►适应新时代

当建模人员和开发人员可以依靠具有指导作用的工具设计或重新设计数据存储时，适应新原则就更为容易，这样他们就可以采用更为动态的模式，以恰当和严谨的方式更快地查询针对NoSQL优化的结构。

注入工具中的最佳实践或Q & A指导可以引导数据库建模师根据对数据查询方式的期望将SQL表及其内部和之间的关系转化为适当的NoSQL集合。例如，在传统关系数据库结构过渡为NoSQL集合中的嵌入式组件的过程中，通过选择性地反标准化，可以为实时解决方案消除长时间查询和慢速性能。

此外，通过利用逆向工程对NoSQL数据库结构进行本地图形可视化，可提高理解现有NoSQL数据库当前状态的能力。因此，可以跟踪这些结构中实际使用的内容，并创建支持系统知识的文档关系（描述仍然存在但不再明确部署在NoSQL环境中的数据关系）。

当作为可对建议NoSQL模型进行URL访问的基于Web的工具提供时，企业也可能获得协作优势，这样数据库管理人员、开发人员、企业架构师、主管甚至业务分析师都可以围绕新模型或现有模型进行对话，因为模型会随着时间的推移而演变。

提高数据访问速度的新方法

(续)

用户可以利用自定义方法自行选择如何在动态数据结构中利用引用（标准化）或嵌入（反标准化）方法。

一般来说，处理集合中较少的对象或实体时最好使用嵌入方法。当然，这取决于查询的复杂性。

当嵌入的实体数量过大时，查询可能变得更加迟缓，而速度优势也可能被削弱。

治理需要更稳固的基础

几年前，Gartner研究副总裁Nick Heudecker对NoSQL的采用者进行了一次非正式调查，发现受访者中数据库管理员仅占**5.5%**。

他在一篇**博文**中指出，“应用程序开发人员现在可能正在从NoSQL中获得他们想要的东西，但是裁撤初级数据主管将导致较大规模企业面临数据质量和信息管理方面的长期挑战。”

这并不是Gartner第一次对NoSQL领域治理表示担忧。在报告“**Does Your NoSQL DBMS Result in Information Governance Debt**”（NoSQL DBMS是否导致信息治理债务？）中，Gartner指出，对于信息管理领导者来说，了解NoSQL部署会通过哪些方式产生某些问题很重要。

第3节

▶将治理引入NoSQL空间

尽管如此，在利用最新NoSQL数据库技术的同时尝试维持底层数据的完整性、质量和治理可能仍然颇具挑战性。

逻辑数据模型中，实体关系图（ERD）提供不同数据类别及其相互关联方式的图片，因此这类模型为数据治理提供了强有力的支持。数据库管理员、数据架构师和数据主管知道逻辑ERD对跨域建模和管理的重要性。

但是逻辑建模并没有得到NoSQL开发人员广泛接受，因为他们喜欢物理模型，所以很难将NoSQL数据以原生形式无缝集成到更大的数据世界中。

在这方面，我们需要做出改变，以获得对逻辑建模的支持，从而便于引用大量数据源并最终将其映射

回单组企业数据元素，同时为安全性、隐私和其他要求提供适当的治理。

通过使用适当的数据治理工具，我们可以将NoSQL数据库的逻辑建模与其物理建模和呈现关联起来。

例如，一家公司可能会使用此类工具促使NoSQL开发人员从由逻辑ERD抽取对象组成的现有数据词汇表中进行选择，然后再创建新集合或物理对象，从而实现更全面的治理。

需要考虑的其他治理方面就是跨业务和技术角色提供数据模型方面的协作所带来的好处。通过这种方式，个人可以公开表达不同的观点，因而有可能发现模型中其他职能部门人员可能看不到的风险，从而通过流程减轻风险。

第3节 (续)

报告的摘要提出了以下观点：

- “灵活的模式数据库支持对未纳入现有信息治理实践的结构化和非结构化数据进行存储。
- NoSQL数据库将业务逻辑从数据存储转移到应用层，造成信息治理复杂化。
- NoSQL DBMS在大多数环境中可以进行累加，因而增加了数据移动和存储位置的数量。”

NoSQL数据治理在其他几个方面受到了阻碍。

在对基于Web和云的应用进行敏捷开发的快速发展组织中，治理流程和领导能力并不总是能够跟上步伐。

还有一些工具能够将不同的数据源映射到单个逻辑企业模型中，以改进模型表达的治理和协作。

行业领导者知道这种事态不可能继续下去，因此他们之中思想最具前瞻性的人正在采取措施解决这些问题。

数据治理市场：呈上升势头

▶资料来源：MarketsandMarkets

2016

8.632亿美元

2021

22.347亿美元



erwin DM NoSQL一瞥

随着NoSQL在每个转变成数据驱动型企业的公司中站稳脚跟，业界必须对公司转换关系模型、审查和展示现有NoSQL数据库结构以及推进新项目的要求做出回应，而所有这些都助于加强治理实践。

为此，数据建模行业的领导者erwin正在扩展其erwin Data Modeler，以纳入erwin DM NoSQL。

与erwin Data Modeler一样，erwin DM NoSQL体现了该公司的Any²(“任意”方式 - 任何储存位置, 任何数据)数据管理方法 - 能够将“任何储存位置”的“任何数据”可视化，以便更好地理解并做出决策。

简而言之，erwin DM NoSQL是第一个支持NoSQL数据库平台领导者MongoDB的基于Web的本地部署企业级建模解决方案。将来还会增加其他NoSQL数据库支持。

建立在erwin Data Modeler的坚实基础之上

*erwin Data Modeler*连续三年（**2014年、2015年和2016年**）被DBTA杂志读者评为**第一名**。

MONGODB在数据库管理系统中最受欢迎

全球有50000名数据管理专业人员使用erwin的业界领先数据建模解决方案。

330 systems in ranking, June 2017

Rank			DBMS	Database Model	Score		
Jun 2017	May 2017	Jun 2016			Jun 2017	May 2017	Jun 2016
1.	1.	1.	Oracle	Relational DBMS	1351.76	-2.35	-97.49
2.	2.	2.	MySQL	Relational DBMS	1345.31	+5.28	-24.03
3.	3.	3.	Microsoft SQL Server	Relational DBMS	1198.97	+14.84	+23.33
4.	4.	5.	PostgreSQL	Relational DBMS	368.54	+2.83	+81.94
5.	5.	4.	MongoDB	Document store	335.00	+3.42	+20.33
6.	6.	6.	DB2	Relational DBMS	187.50	-1.34	-1.07
7.	7.	8.	Microsoft Access	Relational DBMS	126.55	-5.33	+0.32
8.	8.	7.	Cassandra	Wide column store	124.12	+1.01	-7.00
9.	9.	10.	Redis	Key-value store	118.89	+1.44	+14.33
10.	11.	9.	SQLite	Relational DBMS	116.71	-0.64	+9.02

排名第五的MongoDB在三个NoSQL数据库中处于领先地位，因此进入了十大DB引擎榜单，该榜单根据受欢迎程度对数据库管理系统进行排名。

资料来源：DB-Engines Ranking

第4节 (续)

现在，企业将能够支持NoSQL数据建模的三种主要用例，并使用多种高级功能。

具体来说，通过使用erwin DM NoSQL：

- 用户将能够导入物理和逻辑关系模型，并将它们迁移到MongoDB，从而利用图形和树形可视化以及内置引导编辑器和正在申请专利的Query-Optimized Modeling™，根据应用程序和查询需求，自动嵌入、引用或自定义其基于云的集合转换选项。
- 用户将能够对现有的MongoDB数据库进行逆向工程，以推断其结构并获得集合的可视化展示，并将其嵌入数据库中。
- 用户将能够通过从头开始创建MongoDB模型来支持敏捷开发方法，并且可以使用新的集合和数据对模型进行全面更新，以便为新应用程序提供服务。
- 包括所有类别数据专业人员、开发人员和业务同事的团队可以在模型上进行协作，以便创建场景并进行应用程序代码更改。
- 可以对审核后的模型进行正向工程，并且可以将脚本推送到MongoDB实例中，以验证作为动态结构一部分的顶级集合。
- 将支持逻辑数据模型，以实现对于对象、实体、表单和集合的治理以及NoSQL和其他数据的集成。
- 系统知识将通过关系文件保存。

- 用户最终将能够在企业架构和业务流程的上下文中对数据进行建模，从而进一步打造结构合理的数据驱动型企业。

MONGODB：相对稳定的攀升



►资料来源：DB-Engines Ranking

►专为我们今天的工作方式设计

erwin还将erwin DM NoSQL设计成一种SaaS解决方案，作为对MongoDB的补充，同时迎合新一代“云中诞生”的技术人员的需求，使公司能够按需开始使用，而不会产生间接成本。

erwin认为在NoSQL建模流程中，协作是非常重要的，这一点在其许可模式中得到了体现。

每个开发人员许可证都能用于邀请另外五名参与者查看和评论模型，因此可以在敏捷环境中形成由MongoDB集合、服务和应用程序使用者提供的连续反馈回路。这有助于形成更加精妙的流程，从而在数据库上构建应用程序，利用符合业务需求的数据，并确保模型没有违反治理要求。

想要在关系模型和MongoDB部署之间创建更好连接的现有erwin Data Modeler客户，以及其他只是希望更轻松地发现并理解其NoSQL数据的客户，都是erwin DM NoSQL的功能的主要受益者。

如果想要了解更多信息，请[访问 erwin](#)。



关于erwin
erwin®

erwin, Inc.是全球最大的数据建模软件供应商，受到60多个国家50000多名数据专业人士的信赖。

erwin通过集成业务流程建模、企业架构、数据治理和协作解决方案，构建出业界唯一的统一数据管理平台，将任何存储位置的任何数据整合在一起，以改进和壮大公司运营。全球政府机构和领先的金融机构、零售商和医疗保健公司都将erwin数据管理平台作为其任务关键型数据计划的基础平台。erwin总部位于纽约州梅尔维尔。

更多信息，请访问
erwin.com

