

理解大数据

企业架构师的一天

ENTERPRISE MANAGEMENT ASSOCIATES® (EMA™) 白皮书
供 IBM 使用

2017 年 1 月



IT 与数据管理研究、
行业分析和咨询

理解大数据： 企业架构师的一天

目录

简介	3
新颖而又令人兴奋的数据源	3
不断扩大的项目列表	4
多样化平台	4
如何缓解企业架构师的痛苦	5
培训内部员工，而非聘请昂贵的承包商	6
通过可定制应用加快实现进度	6
与业务利益相关方进行沟通 / 协作	6
通过混合数据生态系统强化自助服务架构	7

理解大数据： 企业架构师的一天

简介

企业架构师的生活变得忙碌而艰难。在大数据时代到来之前，企业架构师“唯一”需要担心的是企业数据中心的数据和系统。然而，在过去的十年里，企业利用信息的方式及数据管理平台支持现代数据源信息的方式发生了翻天覆地的变化。

云计算打破了企业数据中心的重重界限，摆脱企业“四堵墙壁”的限制，外部托管应用及创建数据。过去，企业架构师只需专注安全、隐私和控制工作，这种变化无疑为他们带来了大量麻烦。在移动技术的影响下，数据不断流向数据中心外部。维护流向各数据访问点的数据（往往通过平板电脑或移动电话传入）无疑会制造更多的麻烦。移动设备传入的数据不仅引入了新的数据格式，还带来了洪水般的信息，使企业架构师措手不及。

对于希望建立竞争优势的数据驱动型企业而言，系统与数据的格式和位置的改变引发了巨大的变化。企业可通过以下几种方式奠定优势：利用设备传感器日志、社交媒体流和移动设备地理定位信息等新型数据源；创建新项目利用这些新型数据源；建立适当环境通过不同数据管理平台支持相关工作。

企业架构师综合利用经验、专业知识和企业愿景，为企业提供技术基础架构方面的指导和建议。企业架构师会考量业务需求和技术能力，为企业确立未来发展愿景。这些未来战略需要与现有技术基础架构紧密配合，从而实现未来发展。

新颖而又令人兴奋的数据源

经历过去十年的数次转型，企业内部和外部额外托管及创建了一系列数据源，企业架构师务必仔细考量这些数据源。

外部第三方内容数据源（通常通过基于云的提供商提供）有权更改数据结构，而且无需通知使用该信息的下游组织。事件日志和设备传感器信息也会因各自的配置存在一定的可变性，而且通常以 XML 或 JSON 等多等结构格式呈现。客户和公众创建的社交数据源都是文本格式和音频/视频内容。鉴于信息性质的影响，文本和音频/视觉内容难以存储和利用。

另外，传统关系数据源也被卷入大数据变革浪潮，但同样面临一些自身独有的挑战。来自数据中心外部的第三方信息或基于云的企业实现数据越来越多，因而企业架构师必需设法了解如何利用这些信息。



不断扩大的项目列表

大数据信息规模不断扩大，企业架构师需要设计、实施和管理的项目总数也相应增加。随着数据驱动型企业不断整合大数据源，势必会发现新的数据用途。在近期的 EMA/9sight 最终用户大数据研究中，受访者表示在过去几年见证了数据的大规模增长。

2013 年，近 1/3 的企业表示正在开展一个大数据项目。一个项目意味着一套要求及一项实施和维护任务。而今，这一数值不足 1/5。

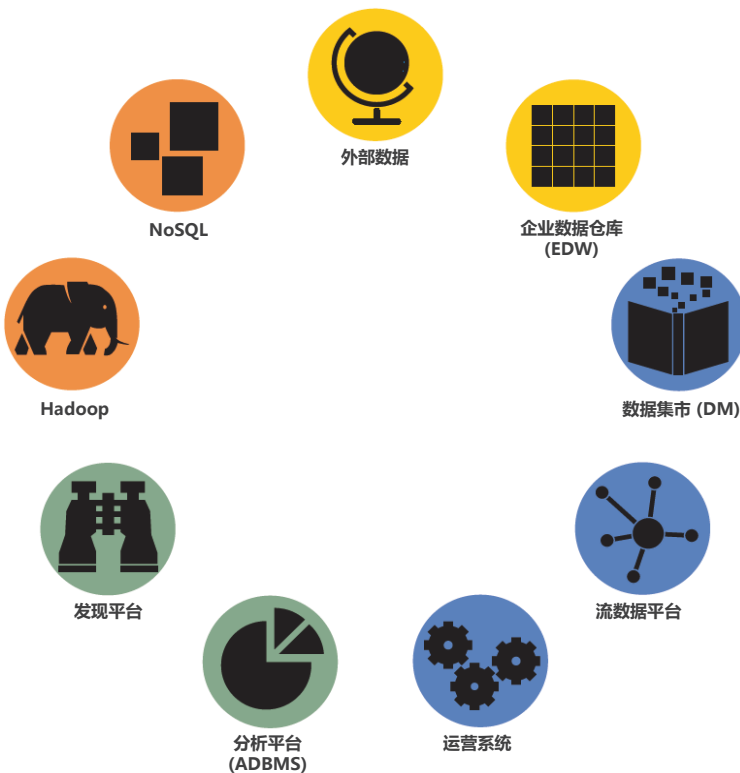
相应地，正在实施多个大数据项目的企业数量翻了一番。2013 年，仅有 32% 的受访者表示实施了三个或三个以上项目。而今，这一数值增长到近 61%。每家企业的项目增量至少相当于收集和实施需求量增长三倍，支持这一增长的数据使用者人数同样增长三倍。

2 倍

2016 年开展**三个或三个以上大数据项目**的企业相当于 2013 年的**两倍**。

多样化平台

随着数据源和项目数量的增加，企业架构师发现支持相关大数据计划的数据管理平台数量同样大幅飙升。在过去的五年中，参与 EMA/9sight 研究的小组受访者一致表示，他们积极利用各种平台提供大数据支持。此类平台不仅限于 Hadoop 或 NoSQL，还包括传统关系型数据库（如企业数据仓库和数据集市）和运营支持平台（如 ERP 和 SCM 等企业应用）。Cassandra 和 MongoDB 都是典型的 NoSQL 平台。流式分析平台可分析从应用和 IoT 设备流出的数据，我们同样将这类平台视为不断发展的生态系统的一部分。数据发现平台介于传统和非传统数据管理范例之间，允许用户探索关系型数据源和多结构数据源。



理解大数据： 企业架构师的一天

在各种不同数据管理平台的支持下，企业架构师可以混搭平台，满足混合业务需求。Hadoop 和 / 或 NoSQL 中存储的事件级数据与运营系统或企业数据仓库中的参考数据相互融合发力。来自第三方的井井有条的客户人口统计数据或企业统计结构数据对现有的 EDW 信息进行了补充，同时为 Hadoop 和 NoSQL 数据提供了额外上下文。在 EMA/9sight 大数据最终用户研究中，66% 的受访者表示正在使用其中 2-6 个平台支持其大数据计划。EMA 将此概念称为混合数据生态系统 (HDE)。

如何缓解企业架构师的痛苦

面对大数据时代带来的种种革命性变化，企业及其企业架构师有两条路可选。第一条路是被动等待大数据趋势到来，意味着企业将一直等到别无选择再做出改变。技术环境和实施策略架构调整采用手动方式且进展缓慢。另外，被动反应型企业几乎只关注技术层面，不愿与业务利益相关方合作。这无疑会限制新型大数据源及现有业务和技术组件发挥价值。

第二条路是企业主动出击，积极做出响应，利用转型机会打造实用的企业架构，包括采用高效方法快速实施大数据项目，同时保持灵活性，以便迎接未来变革。这涉及到对企业文化进行变革，加强内部能力培养，而非走出企业寻找援助，确保实施人员快速找到解决方案，同时企业还要增进与业务利益相关方之间的关系。

66%

的企业在大数据环境中使用
2-6 种不同平台。

数据运营

名词

da·ta op·er·a·tions |

'dā-tā ä-pä-'rā-shāns |

类似于系统管理或网络运营，数据运营涉及组织内部数据流动、存储和保留方面的管理。然而，数据运营不一定牵涉数据战略的使用；仅仅是指维护和恢复问题数据。类似于按一组流程行事（正确做事）但不一定以愿景为指导（做正确的事）的管理战略。

#1

最佳项目实施策略： 可定制应用

通过可定制应用加快实现进度

当主动型企业与企业架构师共同推进大数据计划时，势必希望寻找既能缩短实现时间又能适应未来环境调整的解决方案。在过去的两次调查中，EMA/9sight 小组受访者表示，他们最热衷于利用具有“即时”实施配置的第三方应用（而不是利用定制应用代码）来实施未来调整。

选择此类实施方案后，企业架构师可以灵活地加快大数据项目进度，伴随企业的不断发展深化企业经验。例如，可配置的高级分析大数据解决方案允许在短期内实施并利用初始安装。随着企业对分析模型及其业务应用的了解不断深入，变革在所难免。倘若解决方案可配置且无需定制编码，内部人员将可以灵活迅速地对配置做出“调整”。

培训内部员工，而非聘请昂贵的承包商

鉴于能够轻松配置与大数据项目相关的软件，企业纷纷选择培养内部员工（2016 年占调查小组成员的 18.2%），而不是求助外部资源。企业可以汲取 Hadoop、NoSQL 及其他大数据技术方面的外部专业知识来满足人员配置需求。然而，倘若内部培养大数据人才，那么对公司目标具有深入了解的内部员工将可以发挥自身现有的专业知识和新技能。这样，将可以为企业架构师提供最佳员工组合，促进开展大数据计划，感受到企业对这些活动的投入和重视，而不是一味采用孤立的业务和技术知识。

18%

近 1/5 的企业通过有机员工发展计划
来培养大数据项目人才

3 连续 3 年，业务利益 相关方成为大数据项目 的首要 消费者

与业务利益相关方进行沟通 / 协作

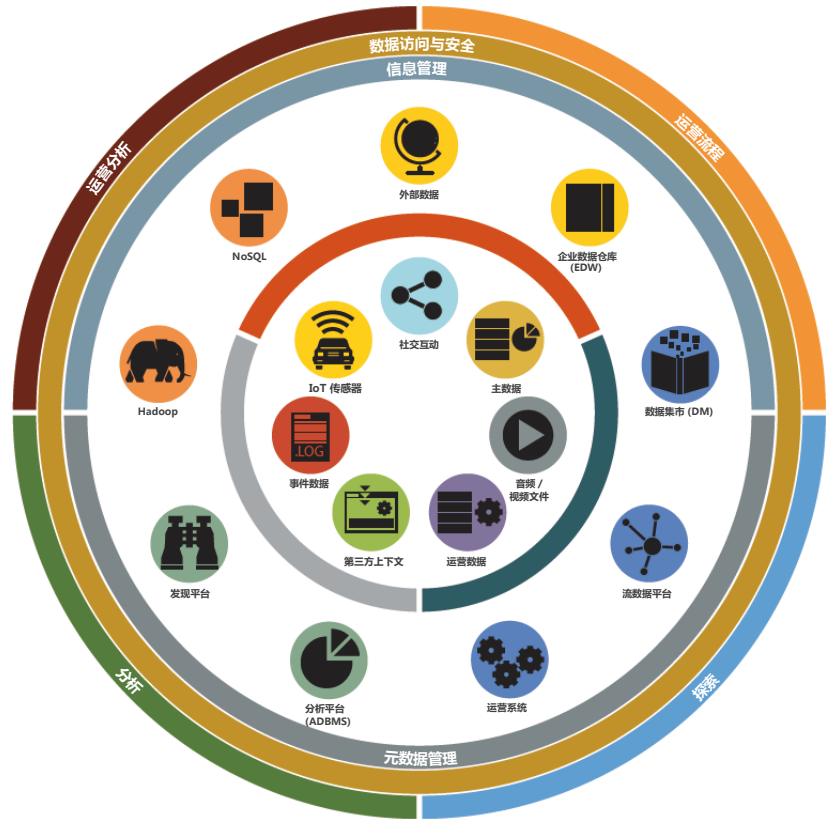
参与大数据项目的内部人员绝不仅限于开发人员。在连续三次调查中，EMA/9sight 小组受访者指出，业务利益相关方（非技术专家）是大数据项目成果和信息的主要使用者。从这一角度而言，主动型企业为企业架构师提供了携手业务利益相关方开展合作的有力工具。这样就可以与运营、人力资源和财务团队等利益相关方更有效地协调一致。

理解大数据： 企业架构师的一天

协作也有利于共同分担若干大数据项目要素的责任。在采取内部人员培养措施后，企业将可以培养技术和业务人才，定义、配置及维护大数据项目环境。

通过混合数据生态系统强化自助服务架构

在精通大数据源的业务利益相关方和优秀的大数据实现技术专家的帮助下，企业架构师将迎来重大机遇。他们可以联合上述两个群体开创新颖独特的解决方案。这种伙伴关系为实施与大数据计划相关的自助服务项目奠定了基础。企业架构师可以面向这些领域提供灵活的大数据框架，包括元数据管理、处理和数据访问工具，以及数据可视化和案例分享工具。



自助服务环境支持业务用户访问并使用企业数据，即使没有数据管理、统计分析、商业智能或数据挖掘学科背景也无妨。数据使用者可以根据自身需求和分析结果做出决策，这样企业的信息技术团队就能专注打造能够赋能数据使用者的战略环境，不再将 IT 工作重点放在战术任务、手动实现方法以及为抓住商机而提升快速响应能力方面。

理解大数据： 企业架构师的一天

业务领域的数据使用者可以管理自身追求，继而通过数据发现或探索活动的形式寻找大数据源的新用途。比如，开展分析项目，通过成本管理活动来调整成本；开展高级分析项目，确定高收益客户的一系列新属性。

这不同于“电子表格数据集市”或非托管影子 IT 项目，并非一次性实现项目。相反，这是从战略角度面向业务利益相关方提供的接受监管的环境，以便业务利益相关方利用公司资产满足自身需求。这种环境可能以新大数据源和平台的形式呈现，能够捕获可供企业使用的高价值元数据。

企业架构师会创建大数据自助服务功能。通过设计和实现采用可配置软件组件的应用环境，让技术专家能够发展技能，积极支持技术实现团队与各业务部门之间的互动，企业架构师将可营造舒适环境，确保业务利益相关方可以按“业务发展速度”（而非实现积压速度）满足要求。企业则可专注研究如何充分利用新的大数据资源，无需等待战术 IT 工作流程，同时维护实现组件，以便面向整个企业推广应用并持续开发。

关于 Enterprise Management Associates, Inc.

Enterprise Management Associates (EMA) 成立于 1996 年，是顶尖的行业分析公司，提供有关 IT 和数据管理技术的全面深入见解。EMA 分析师综合利用独有的实践经验、对行业最佳实践的见解、对当前和计划的供应商解决方案的深入了解，帮助 EMA 客户实现他们的目标。要深入了解 EMA 面向企业业务用户、IT 专业人士及 IT 供应商提供的研究、分析和咨询服务，请访问 www.enterprisemanagement.com 或 blogs.enterprisemanagement.com。您也可以在 Twitter、Facebook 或 LinkedIn 上关注 EMA。

未经 Enterprise Management Associates, Inc 事先书面许可，不得对本报告的全部或部分内容进行复制、拷贝、存储到检索系统或再传输。此处的所有观点意见以及预估都是截止本文发布时我们自身的判断，可能会随时更改，恕不另行通知。本文档中提到的产品名称可能是其各自所有企业的商标和 / 或注册商标。“EMA”和“Enterprise Management Associates”是 Enterprise Management Associates, Inc. 在美国及其他国家 / 地区的商标。

©2017 Enterprise Management Associates, Inc. 保留所有权利。EMA™、ENTERPRISE MANAGEMENT ASSOCIATES® 和莫比乌斯带符号是 Enterprise Management Associates, Inc. 的注册商标或普通法商标。

公司总部：

1995 North 57th Court, Suite 120
Boulder, CO 80301
电话：+1 303.543.9500
传真：+1 303.543.7687
www.enterprisemanagement.com
3515.012617