

# 数据治理和数字化转型



国际数据管理协会中国分会（DAMA China）主席 汪广盛

## 问大家一个问题

---

如果你是国家领导人，对国民经济而言，你认为哪个行业更重要：

1. 制造业
2. 房地产

## 再问大家一个问题

---

就制造业而言，我们现在是否仍然有人口红利 – 劳动力成本的优势：

1. 有
2. 没有

# 为什么要数字化转型？



数字化

- 开发周期：9年→4.5年
- 成本降低：25%
- 100%整机数字化设计



数字化

- 客户满意度提高30%以上
- 人均产值增长20%
- 资产利用率提升8%



数字化

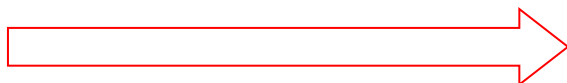
- 产能提高了8倍
- 每天采集数据5000万个
- 产品合格率达99.9988%

# 为什么要数字化转型?

数字化转型的核心是，**提高制造业的水平**



手动式



半自动



全数控



数控织机，由原来3-4小时/毛衣，变为40分钟/毛衣，1个工人操作5-10台机器

## 再问个问题

---

那么好的事情，为什么不转呢？

75%的数据相关的项目都失败了！

**Businesses believe that around 29% of their data is inaccurate in some way.** This statistic has remained unchanged now for several years and points to the fact that under IT control, the basic foundations needed for today's digital business aren't in place.

# 数据治理和DAMA



# 关于国际数据管理协会

## □ DAMA International （国际数据管理协会）

DAMA International成立于1980年，是一个由技术和业务专业人员组成的国际性数据管理专业协会，作为一个非营利的机构，独立于任何厂商，旨在世界范围内推广并促进数据管理领域的概念和最佳实践，为数字经济打下理论和实践基础。全球会员近万人，在世界48个国家成立有分会。  
<http://www.dama.org>

## □ DAMA China （国际数据管理协会中国分会）

DAMA China是个非营利性、专注数据管理的专业组织。旨在交流国际、国内在数据管理领域中的最新进展，共享业界的实践、经验和成果，促进我国数字化水平的不断提高和创新。  
<http://www.dama.org.cn>

### DAMA 全球成员单位

Australia  
Austria (Oesterreich)  
Belgium  
Canada  
China  
Denmark  
Finland  
Germany (Deutschland)  
India  
Luxembourg  
The Netherlands  
Nigeria  
Norway  
Singapore  
South Africa  
South Korea  
Sweden  
Switzerland (Schweiz)  
United Kingdom

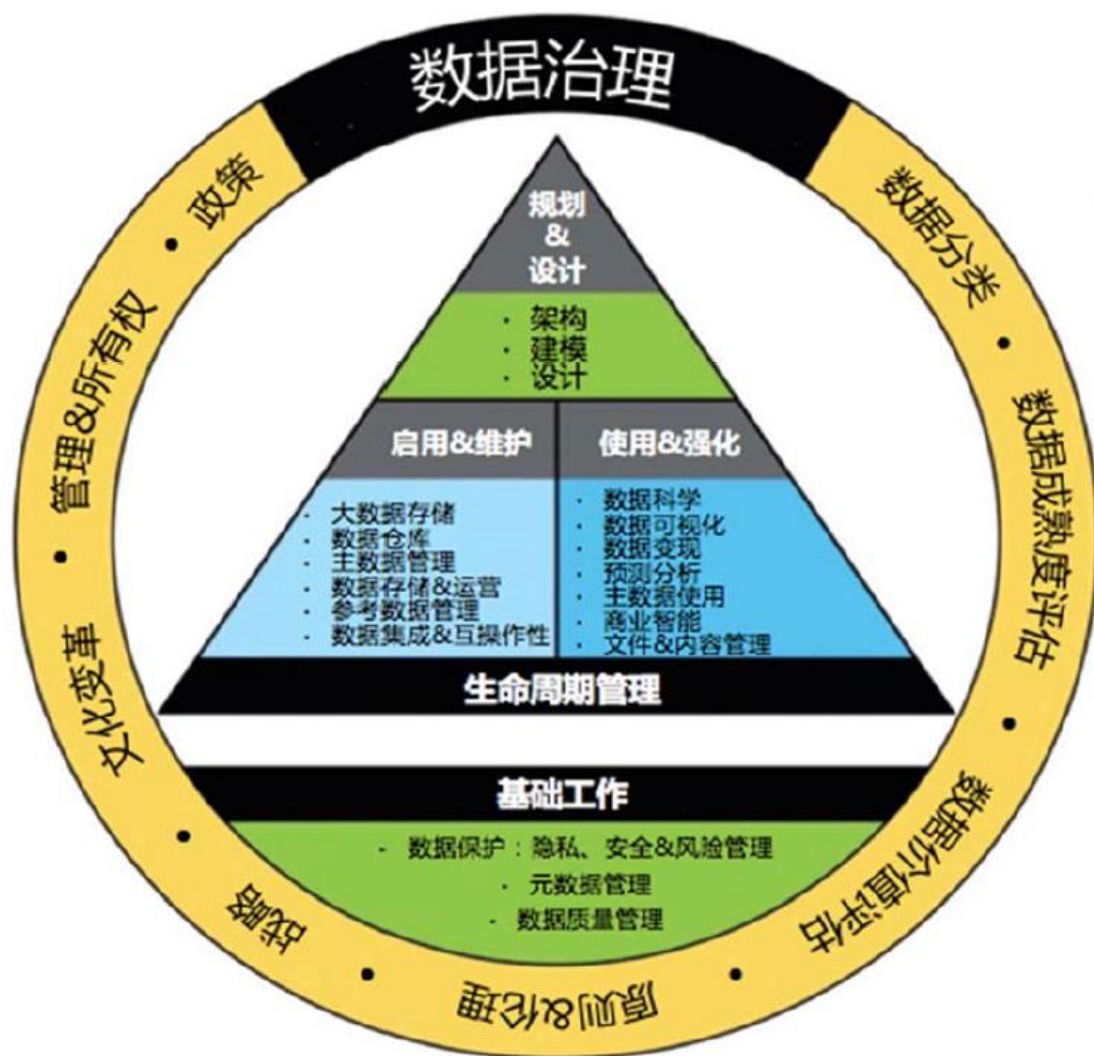
### 美国分会

Arizona  
California  
Colorado  
Delaware  
District of Columbia  
Florida  
Georgia  
Illinois  
Iowa  
Kansas  
Maryland  
Massachusetts  
Michigan  
Minnesota  
Missouri  
Nebraska  
New Jersey  
New York  
North Carolina  
Ohio  
Oregon  
Pennsylvania  
South Carolina  
Tennessee  
Texas  
Virginia  
Washington  
Wisconsin

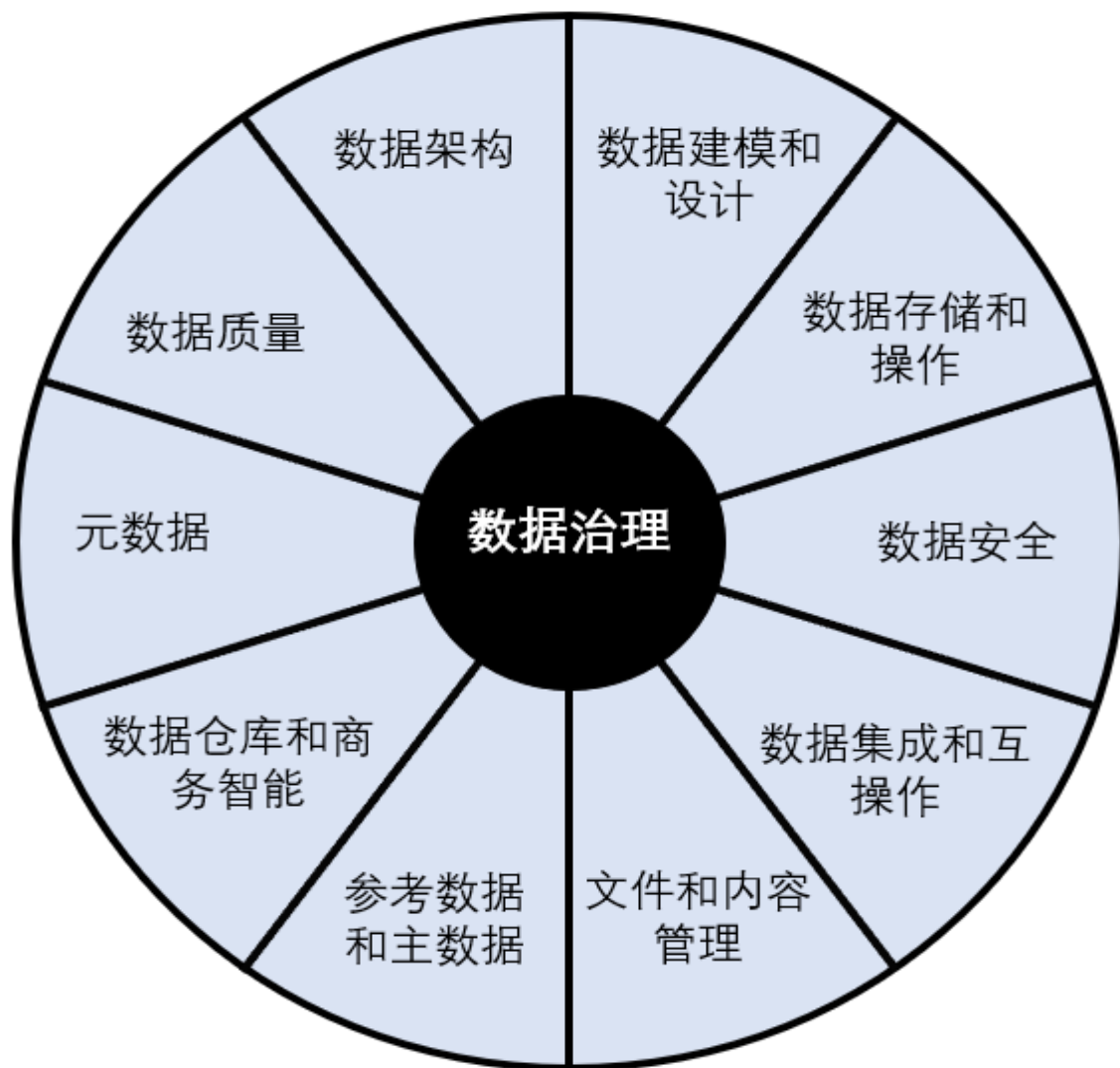




# 数据治理的内容



业务层面



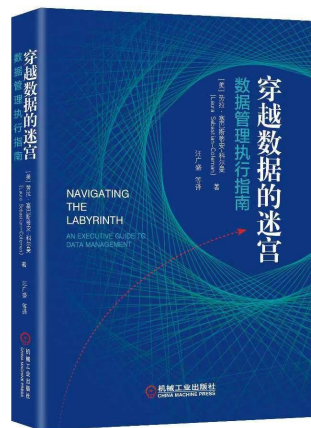
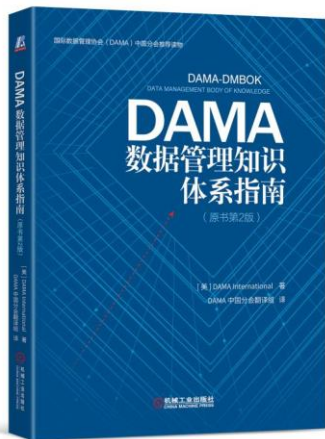
技术层面

<http://www.dama.org.cn>

# DAMA数据管理知识体系和CDMP认证

## □ 《DAMA数据管理知识体系指南》

DAMA集业界数百位专家的经验于一体，编写了数据管理知识体系，共600多页。该体系是业界数据管理最佳实践的结晶，成为从事数据管理工作的经典参考和指南。该指南2.0版已于2020年5月在国内翻译出版。



## □ 数据管理专业人士认证

该证书(Certified Data Management Professional, CDMP) 由国际数据管理协会 (DAMA)统一命题考试和颁发。**国际认证，行业认可。**

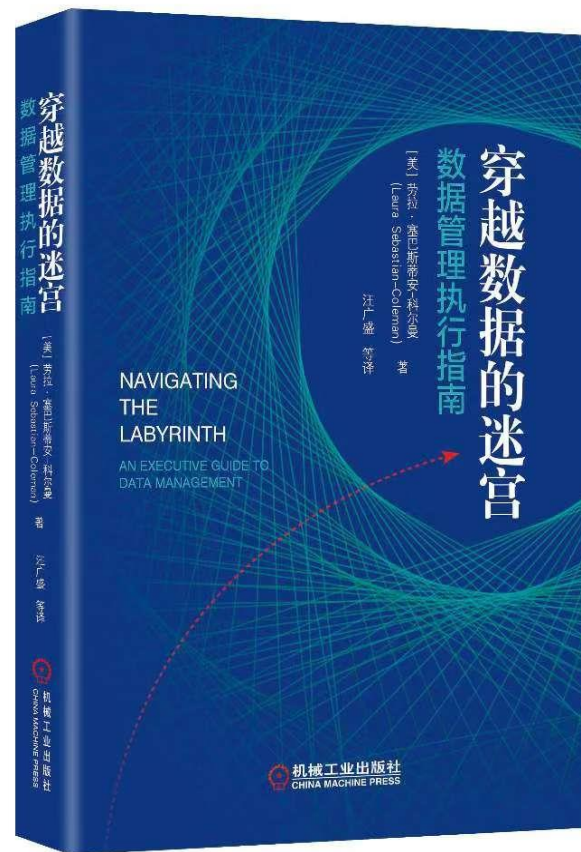


是世界唯一的和最权威的关于数据管理和数字经济的相关认证

# 穿越数据的迷宫

- 作者是Laura Sebastian-Coleman博士，出版于2018年6月
- 机械工业出版社
- 国际数据管理协会（DAMA）中国分会会员志愿翻译
- 专门为管理层和政府相关部门领导而写的Mini版DMBOK2.0
- 共12章

第一章 管理数据的重要性  
第二章 数据管理的挑战  
第三章 DAMA的数据管理原则  
第四章 数据伦理  
第五章 数据治理  
第六章 设计和规划数据生命周期  
第七章 数据赋能和数据维护  
第八章 使用和增强数据  
第九章 数据保护、隐私、安全和风险管理  
第十章 元数据管理  
第十一章 数据质量管理  
第十二章 现在要做什么

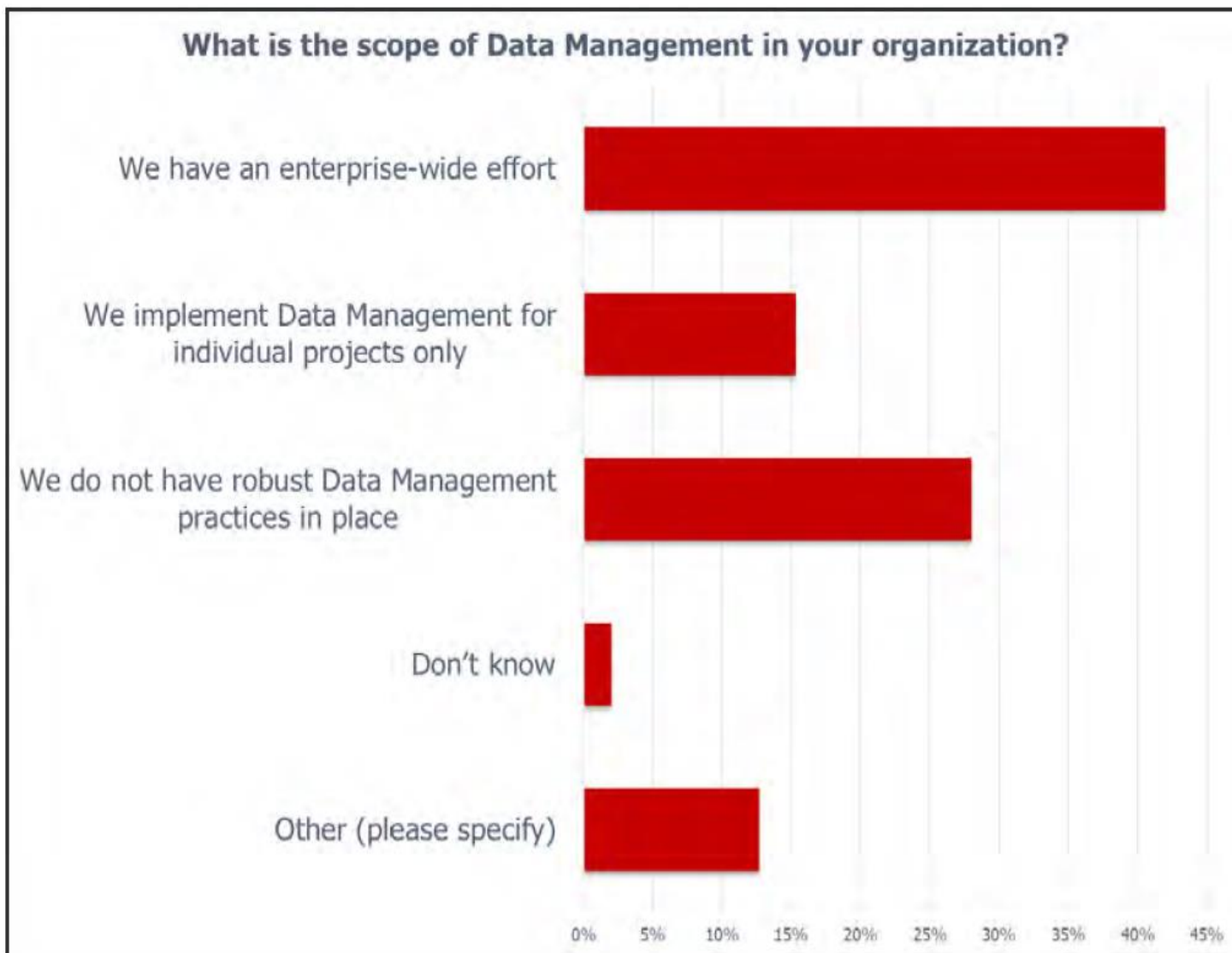


# 数字金融的现状

# 企业级的数字化管理

是否有企业级的数字化管理？

有：42%  
没有：28%

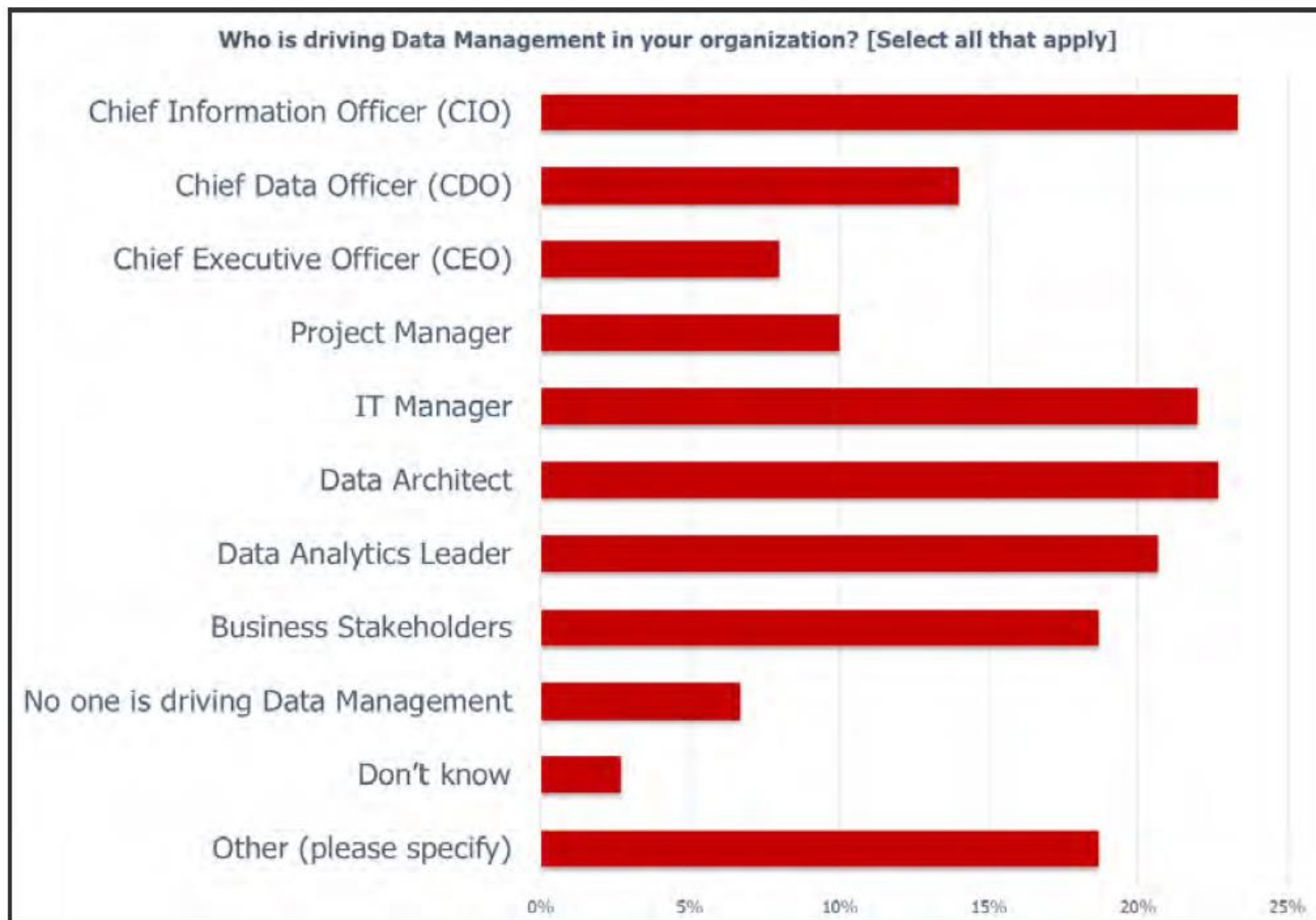




# 谁在推动数据管理工作

谁在推动数据管理？

- 所有人都参与
- C级别领导的支持
- CIO：23.33%
- 数据构架师：22.67%





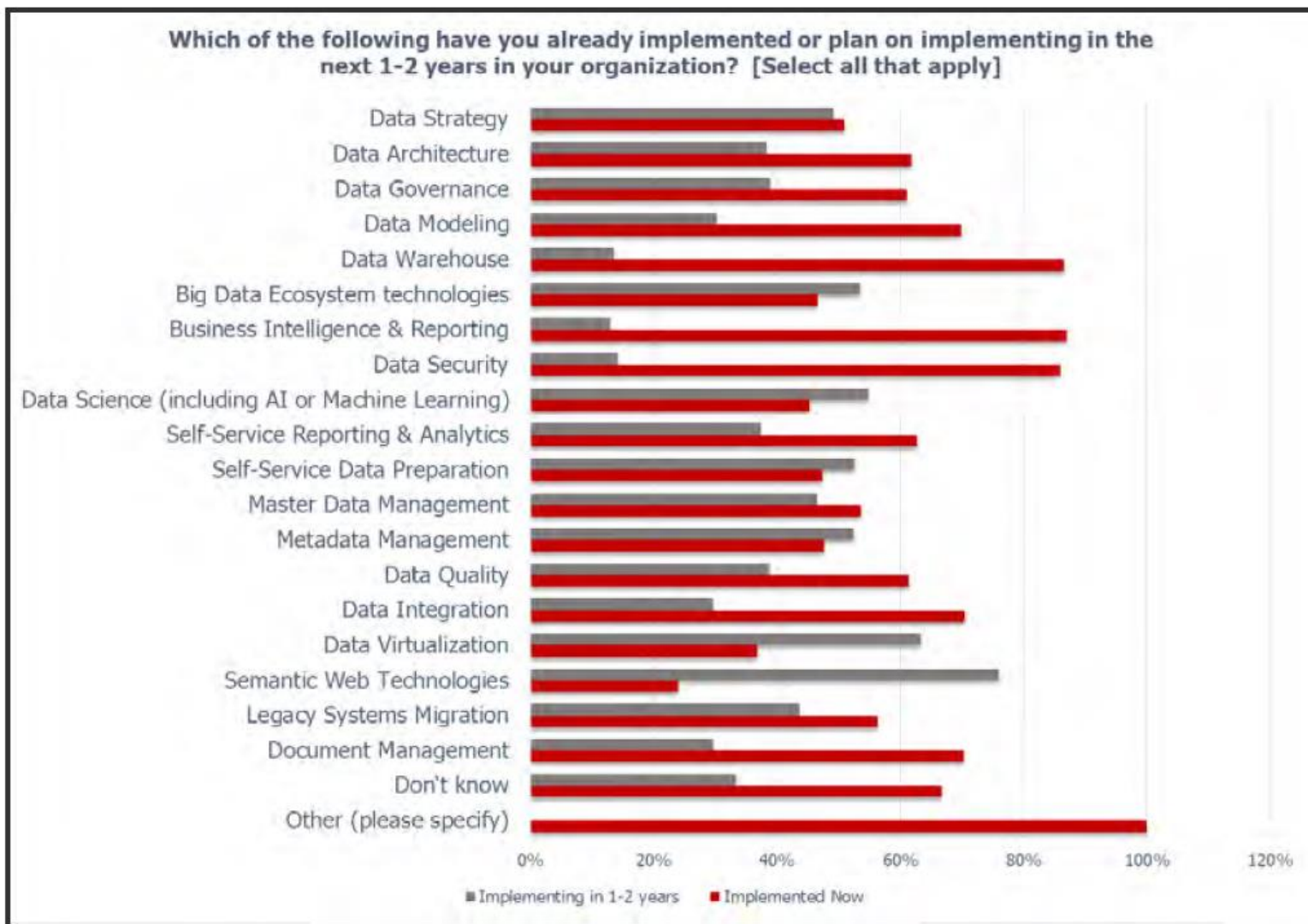
# 数据管理的目的是什么

为什么要做数据管理？

了解公司运营：79.70%

减少风险：66.92%

数字化转型：53.38%



# 数字化转型最大的挑战

---

最大的挑战有哪些？

1. 资源：人力、财力和技术
2. 管理层不够重视
3. 企业构架和企业文化

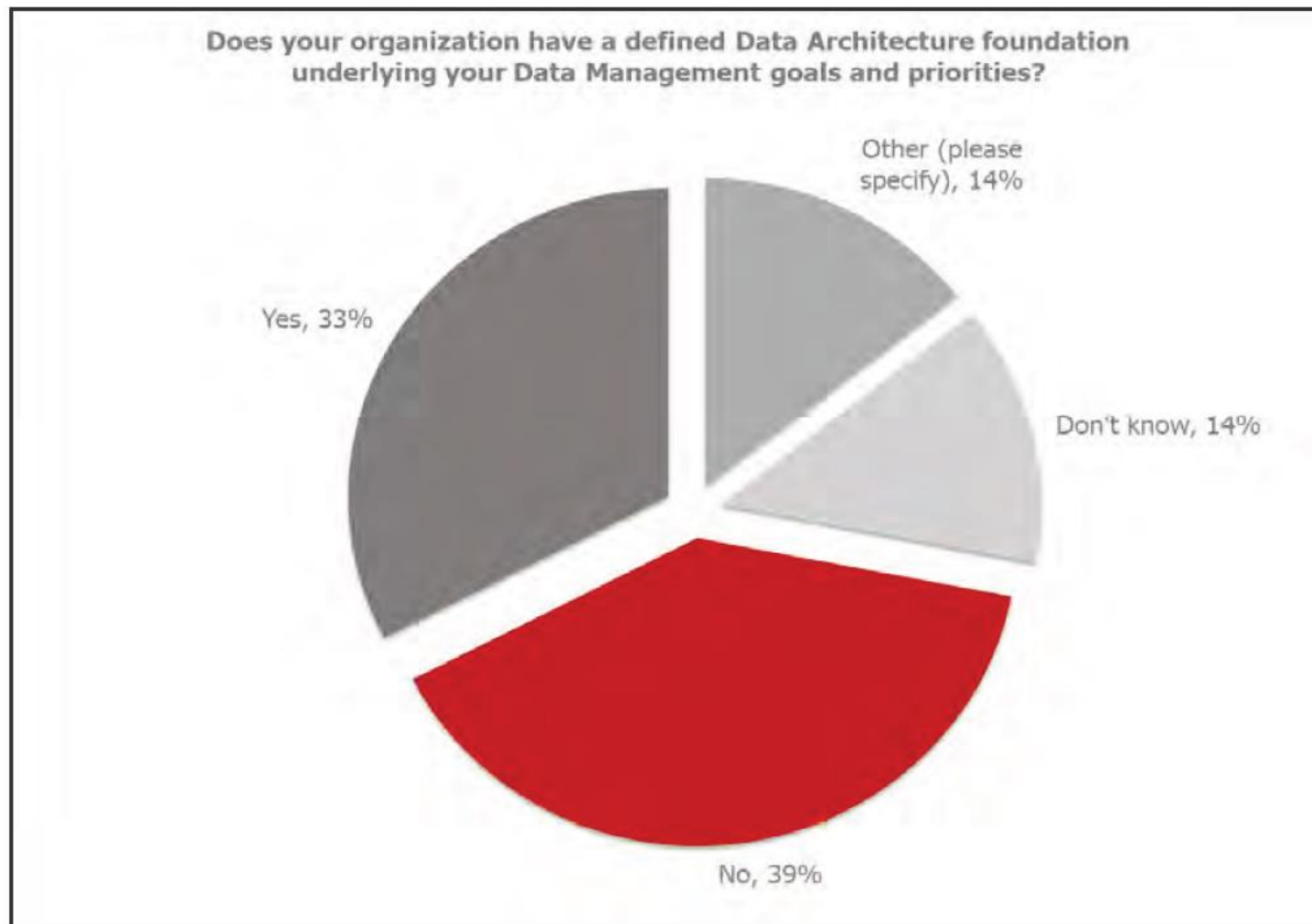


财务报告年度结束后一周就可以出来了

# 数字化转型是否有数据管理的架构

是否有数据构架？

有：33%  
没有：39%

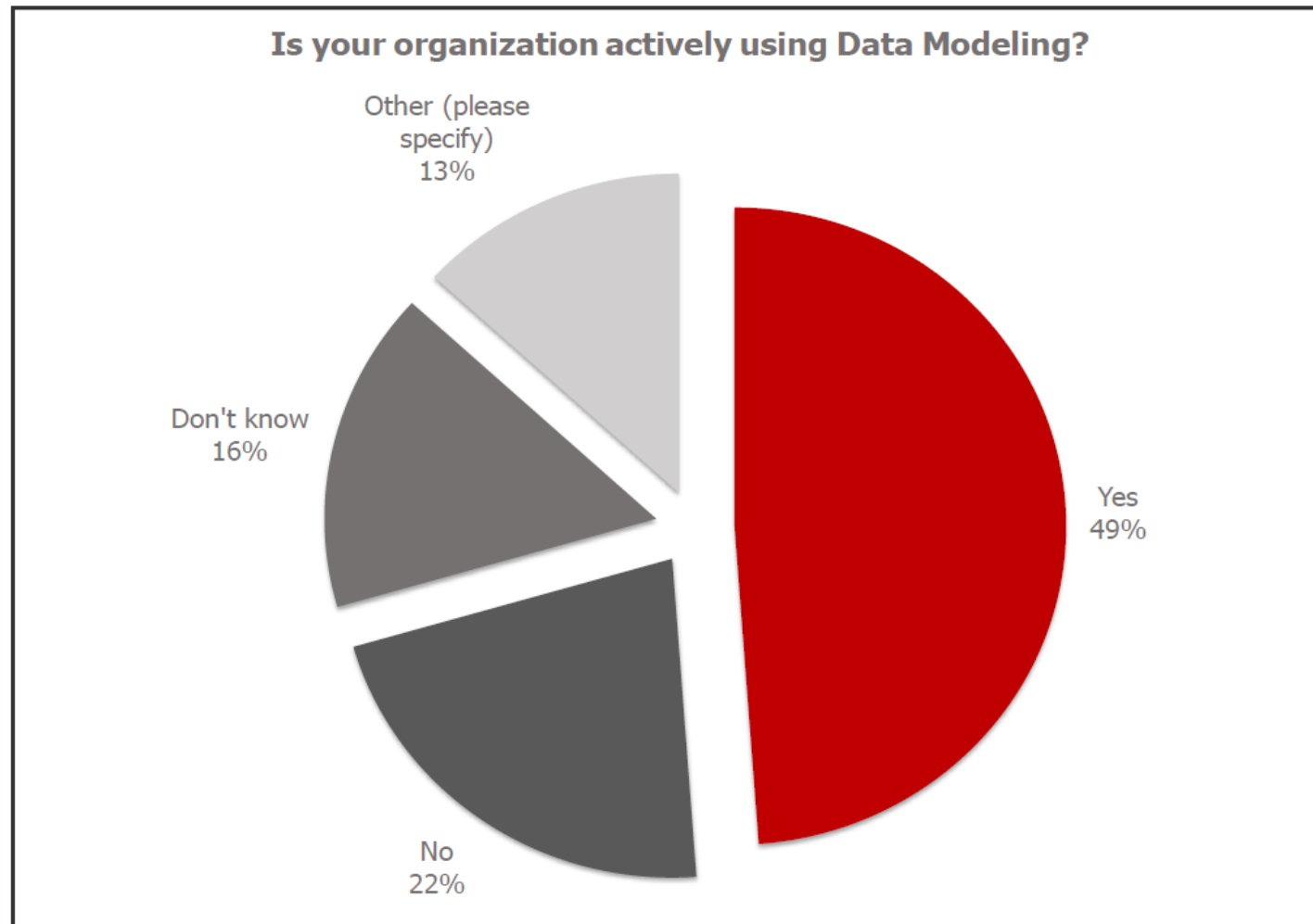


# 数据建模是否需要

是否用数据建模？

有：49%  
没有：22%

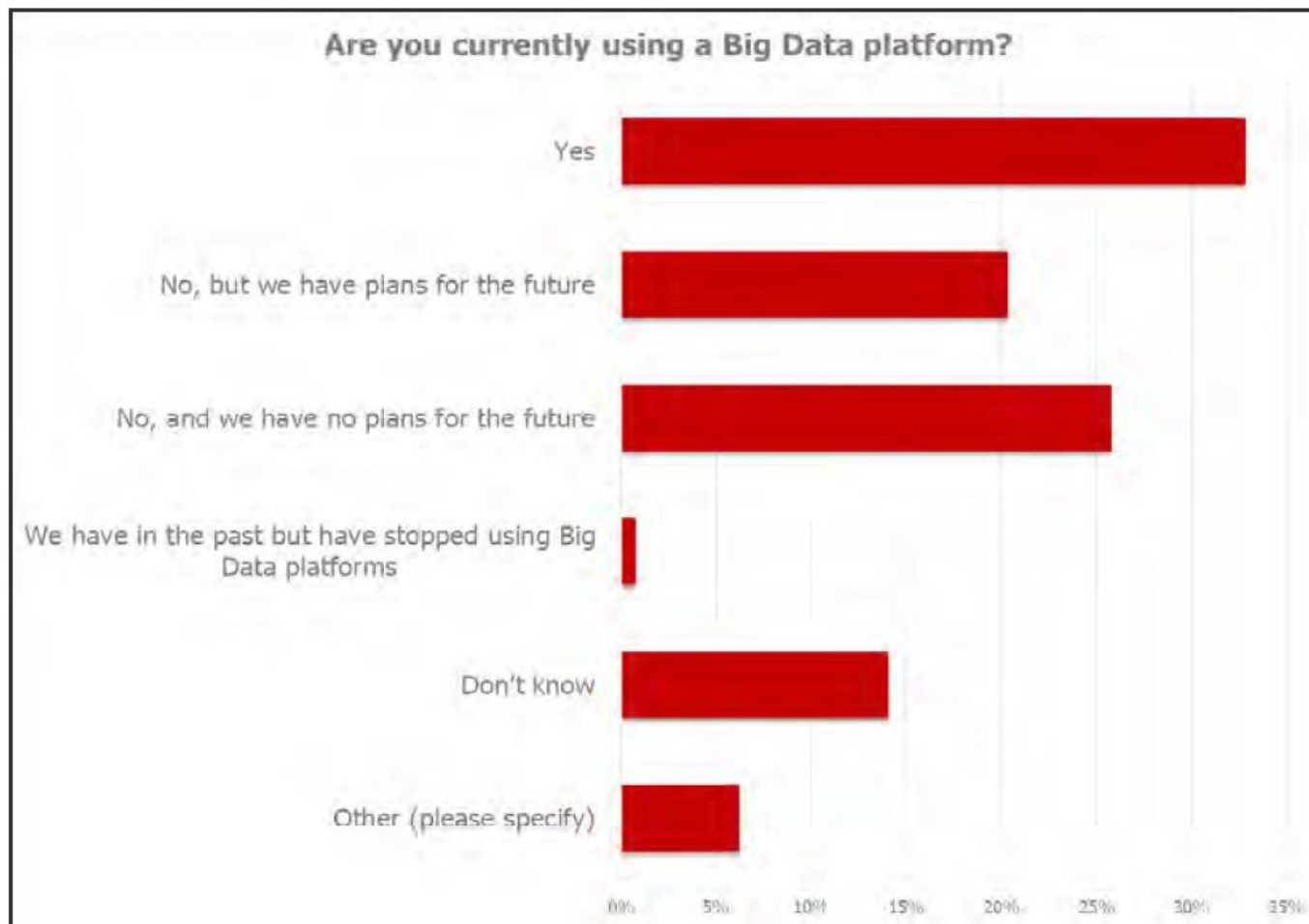
FannieMae有600多人的建模团队



# Hadoop等大数据系统

是否在用大数据系统？

- 是：33%
- 否，暂时以后也不会用：26%
- 否，但以后会用：20%
- 以前有，但现在不用了：1%



# 从数仓到数据湖

是否在用数据湖？

- 是，和DW一起用：20%
- 否：32.26%

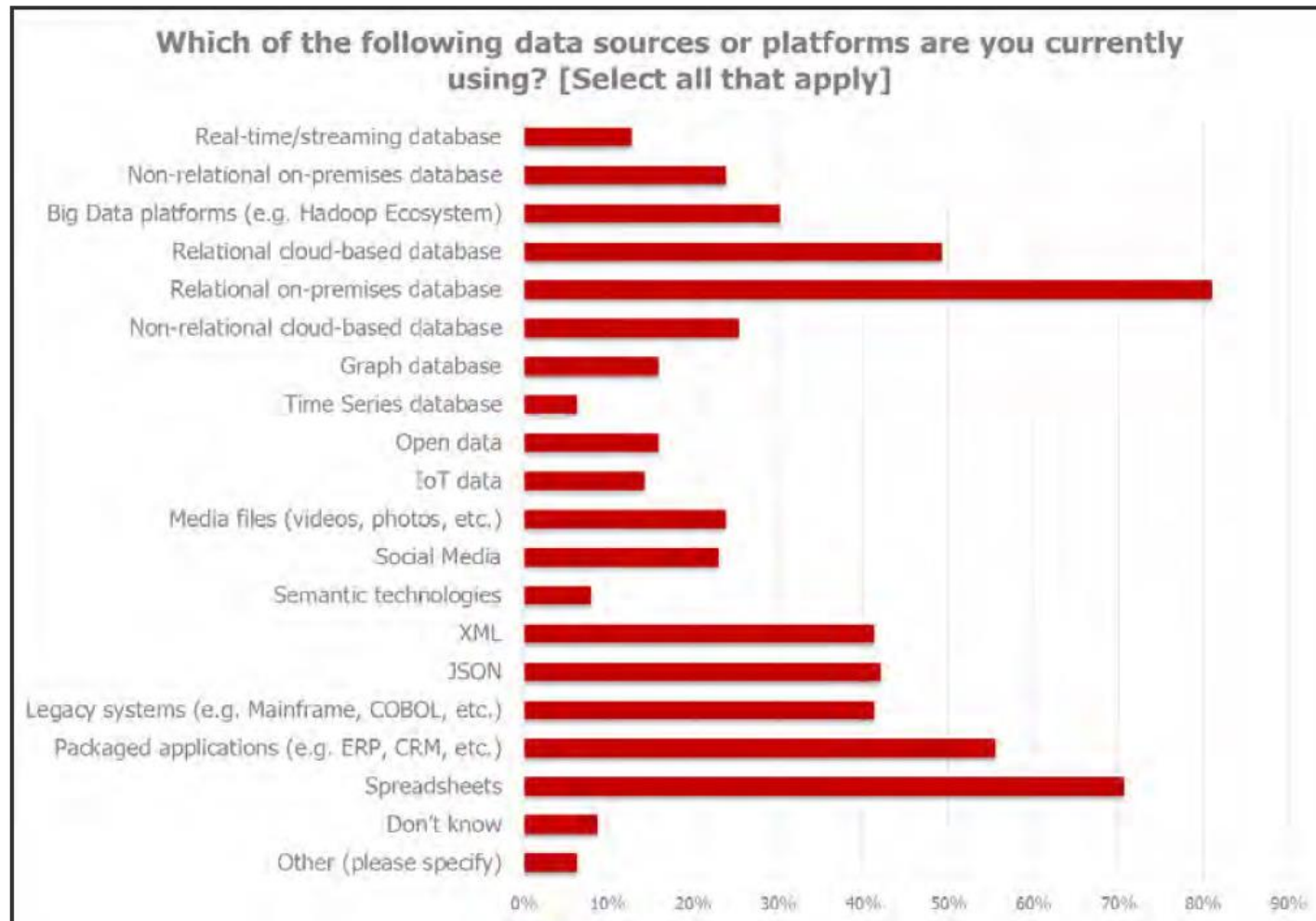




# 数据库和数据存储

## 数据库和存储

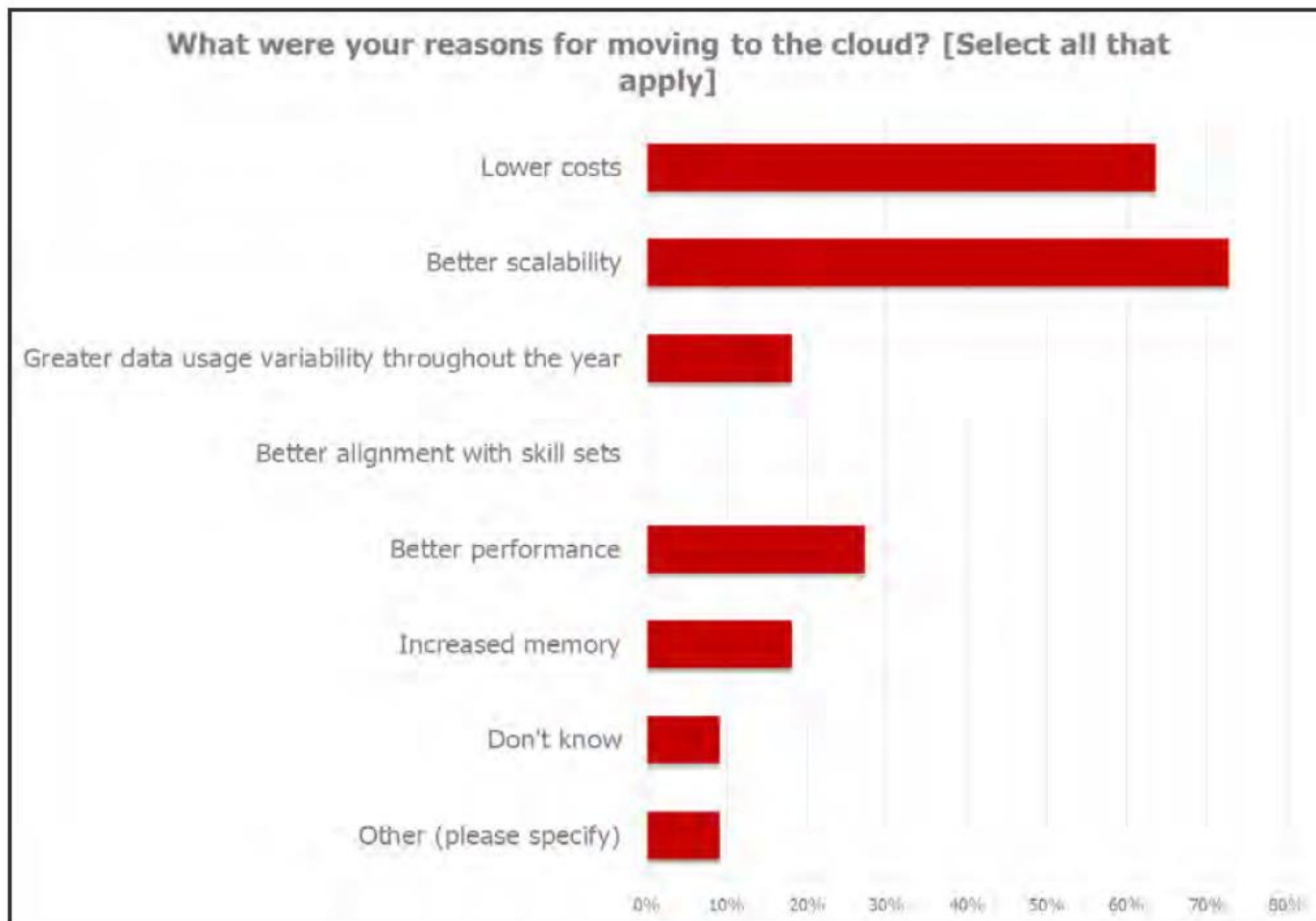
- 传统RDBMS: 80.95%
- 云端RDBMS: 49.21%



# 数据上云端的原因

为什么要上云端?

- 更好的扩展性: 72.73
- 更低的成本: 63.64%



# 数据管理和新技术

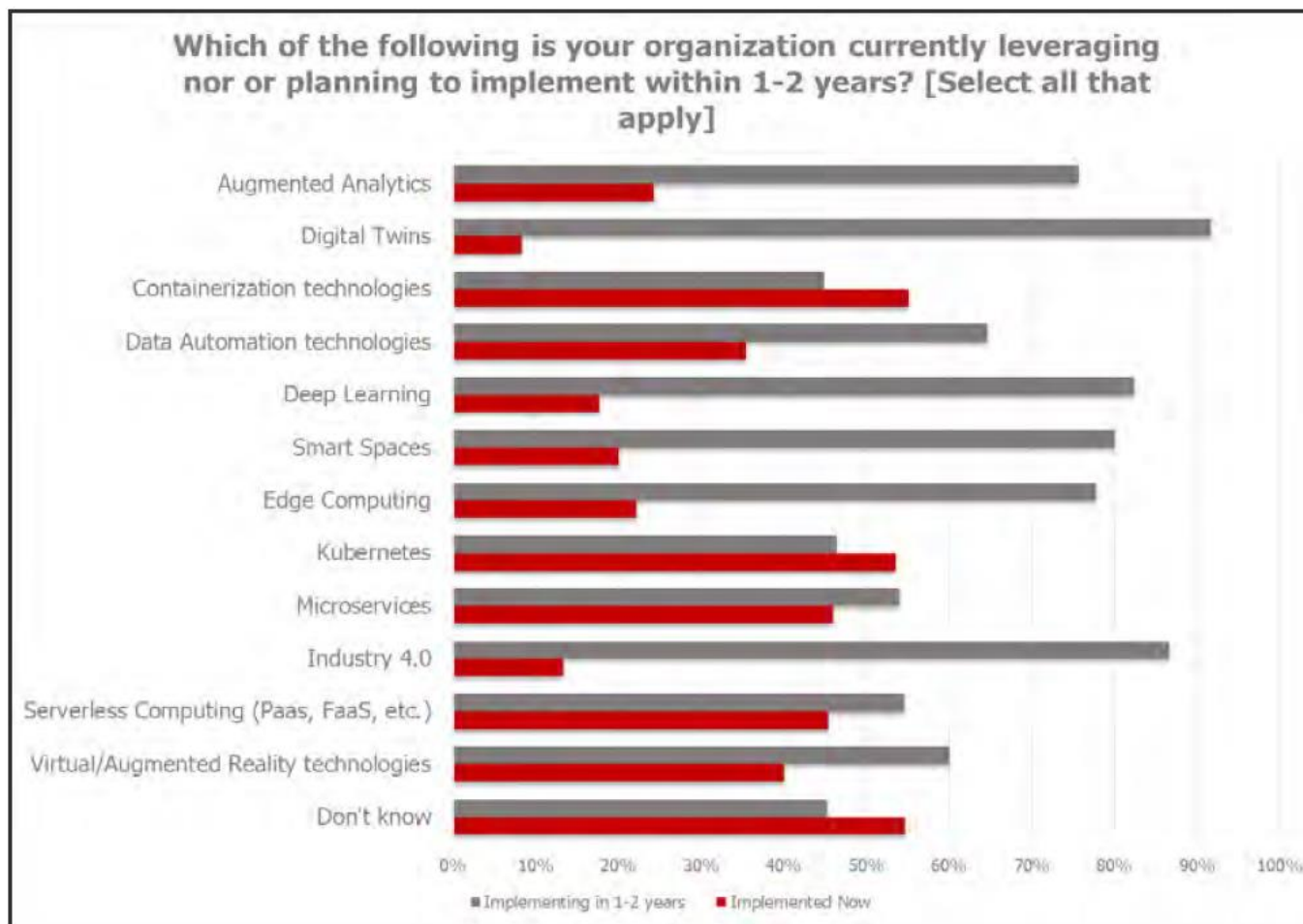
是否已经或者计划在一到两年内  
会用如下新技术

现在:

- Containerized: 55%
- Kubernetes: 54%

以后:

- Digital Twins: 92%
- Industry 4.0: 87%
- Deep Learning: 82%
- Smart Space: 80%



# 数字金融的未来趋势



## 业务：老年人俱乐部 vs. 阅历和伤疤

巴塞尔协议从70年代初发达国家间的一个银行业协调监管机构，到如今成为全球银行业监管标准的制定者，一直得到了全世界主要经济体的支持和尊重，中国亦不例外，2009年中国加入巴塞尔委员会。

2008年金融危机后，巴塞尔协议III提出对银行体系外的“影子银行”一并纳入监管，这块主要是在银行之外从事金融放贷或类放贷融资活动的实体，信托、消费金融公司、支付公司、以及从事放贷业务的互联网公司都属于这一范畴。

其中最核心的一点是什么？银行要满足资本充足率的要求，银行不能以金融创新的名义无限扩大表外业务，但影子银行可以。假如“影子银行”以10亿的资本金运作一个次级贷款池，通过反复加杠杆，最终放大到上千亿规模，杠杆率能达到数百倍。

从这个意义上讲，市场上众多互联网金融放贷机构，如果按银行业标准要求，足额计提风险准备金，有多少家风险会爆表？

如果银行无须计提风险准备金，无须执行资本充足率要求，银行可以将放贷利率提升至法定利率上限，那金融科技公司的所谓大数据优势立刻会化为乌有。

中国银行是1912年成立的，迄今108年；花旗银行是1812年成立的，美国银行是1784年成立的，即便招商银行也是1987年成立的，也有33年了。中国所有放贷的金融科技公司都没有超过10年的从业史。



# 业务：华尔街Fintech六大趋势

---

按照PWC和EY2019年底对美国金融业的调查, 华尔街在今后的 3 到 5 年呈现出如下六大趋势:

1. 提高用户体验
  - ✓ 需求多样化
  - ✓ 个性化服务
2. 大数据和AI:
  - ✓ 数据集成、分析
  - ✓ AI的普及
3. 移动端
  - ✓ 支付
4. 数据安全
  - ✓ 开源系统
5. 主动监管
  - ✓ 自律和监管
6. 业务创新和对传统系统的挑战
  - ✓ 创新“新常态”: 比特币, 区块链; 数字货币

- ❖ 投资风险分析: 3个月到3小时
- ❖ 价格定位: 48小时到20分钟到实时
- ❖ 用户行为分析: 72小时到20分钟到实时
- ❖ 数据自动建模: 每年150个到每年15000个



# 技术：数据管理核心内容没有变

---

## 关于主数据和数据标准及数据质量

我们DAMA群里有次曾经讨论过“主数据是否已死”的问题。有会员认为主数据是特殊时代的历史性产物，现在已经没有意义了。也有人提议说，你说主数据已死，你不是砸人家饭碗吗？毕竟许多搞数据管理的是以此为生的。

首先我觉得这样的讨论很好。可以让我们不时的回过头去看看那些我们已经习以为常的事。不过主数据确实没有死。

就技术本身而言，主数据是为了数据标准和数据质量。当我们看到数据标准还有许多空白，数据质量还有严重的问题，跨部门和跨行业数据治理还没真正开始，我们就知道主数据肯定没有死。当然主数据管理需要有个现代的崭新的视角。另外，到底哪些是主数据？我们用什么标准来选择和判断谁可以是主数据？真正的分歧在这里。

## 关于元数据

数据“自助服务”（Data Self Service）还是需要元数据

数据目录（Data Catalogs）充当最终用户索引，用于搜索通过数据发现编译的数据源和定义。该技术由自动机器学习（标记或扫描数据）提供支持，通常用于确保改进自助服务BI治理。数据目录对于常见的数据管理计划（如提高数据质量和法规遵从性）正变得越来越流行。

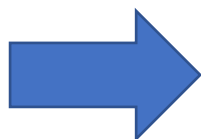
## 关于SQL

这个产生于45年前的语言，今天仍然有强大的生命力。According to Stack Overflow, SQL is the 3rd most popular coding technology. 问题：NoSQL是“No SQL”还是“Not Only SQL”？

# 从合规到生产要素

---

继GDPR和新颁布的《加州消费者隐私法》之后，合规的要求使得数据治理越来越重要。数据治理更多的是为了合规。



2020年4月，《中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见》  
数据作为一种新型生产要素写入文件中，与土地、劳动力、资本、技术等传统要素并列为要素之一。《意见》明确，加快培育数据要素市场，推进政府数据开放共享、提升社会数据资源价值、加强数据资源整合和安全保护。

Forrester reports that data-driven companies grow more than 30% annually on average. Leading companies are capitalizing on data and analytics to widen the gap between them and their competition.

# 增强式数据管理： Augmented Data Management

---

增强式数据管理利用机器学习和人工智能的优势，使一些数据管理任务“自配置、自调整”，这包括数据质量的自动检测和元数据的实时更新管理等。Gartner预测，到2022年底，通过增强式数据管理，数据管理的工作量将减少45%。这些增强现象已经在商业智能和数据分析中看到。这些增强因素可能会使我们在数据管理领域看到一个全新的企业软件类别的发展。

几个问题请大家思考：

- AI是手段还是目的？
- 一个数据治理项目大概要做多久？
- 数字化转型从开始到成功大概要多久？

The goal is to surface critical insights that take less time, less skill, and have less bias. As was brought out in our introduction, data scientists spend over **80%** of their time doing simple mechanical things such as labeling and cleaning their data. Augmented analytics would use AI to do most of the heavy lifting and could even lead to completely autonomous business analytics systems in the future.

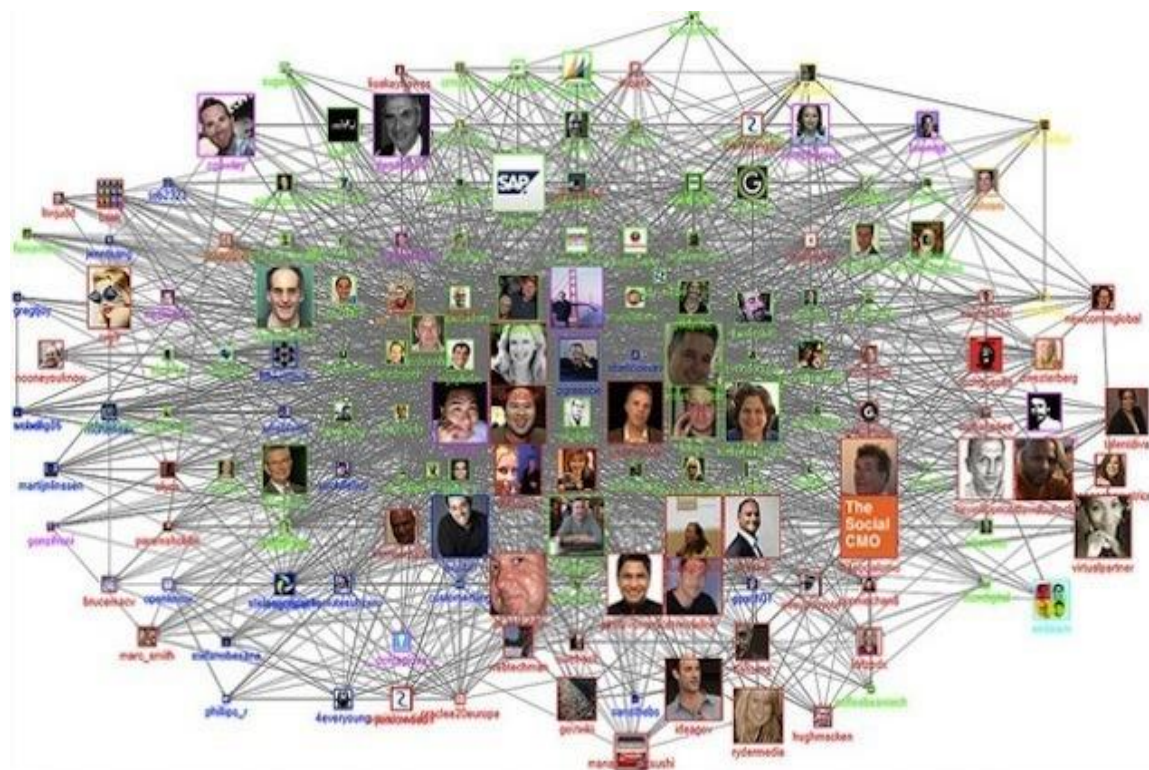
# 图形关系分析：Graph Analytics，和非结构化数据

业务用户提出越来越复杂的问题，我们需要从结构化和非结构化数据中寻找答案，通常是将来自多个应用程序的数据以及越来越多的外部数据混合在一起。使用传统的查询工具或查询语言（如SQL）基本上无法分析这种级别的数据复杂性。

图形分析是一组分析技术，显示实体（如人、地点和事物）如何相互关联。该技术的应用范围从欺诈检测、交通路线优化和社会网络分析到基因组研究。Gartner预测，在未来几年中，图形处理和图形数据库的应用将以每年**100%**的速度增长，以加速数据准备，并实现更复杂和适应性更强的数据科学。

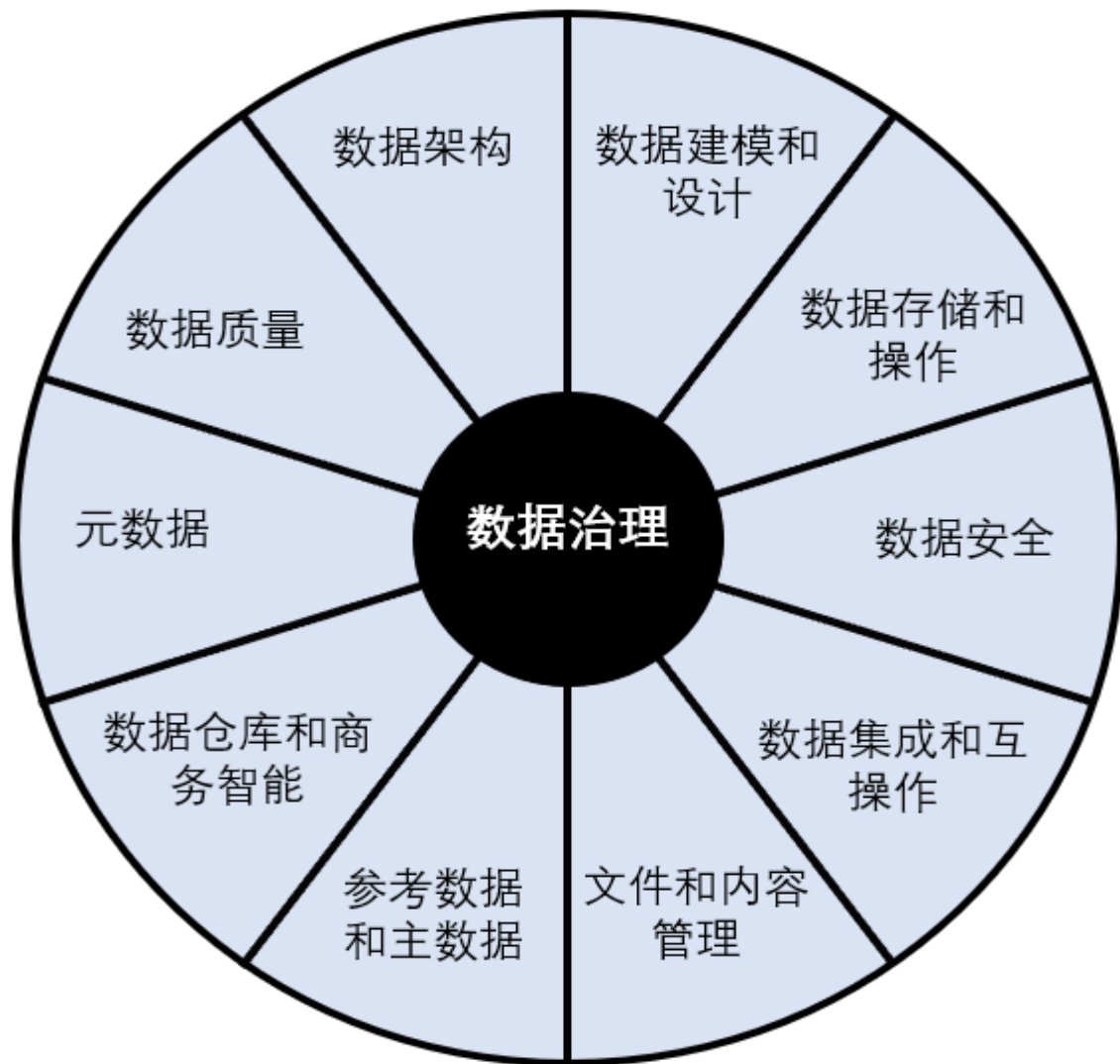
有什么影响：数据架构，BI等

- Path analysis:
- Connectivity analysis:
- Community analysis:
- Centrality analysis:



# 从数据管控到数据资产的管理

- 统一的数据管理平台（Centralized or Single Point Data Management Platform）的使用预计将在今后几年占据主导地位，取代众多行业中的多个分散的解决方案。综合性的统一的数据管控平台可以避免许多重复工作，也更具可扩展性。
- 数据是资产，而不再只是一种资源。数据资产的管理将不同于数据资源的管理。





# 分布式云计算：Distributed Cloud

---

在云计算的第一阶段中，企业已经将其20%的关键应用转移到云计算上。第二阶段将包括移动剩余的80%，考虑到合规性、安全性和地理位置方面的独特要求。这导致了一组复杂的混合和多云环境。

在“混合云”和“多云”环境中进行数据管理将很快成为新的规范。云架构和混合云架构的采用，继续带来对可靠和可扩展数据管理的需求。对云数据管理技术的投资也将扩展到SaaS产品，即数据和数据库即服务产品。然而，主数据管理过程现在可能最适合于云。根据IDC的数据，到2020年，90%的大型企业将从事数据即服务（Data-as-a-Service）的创收业务。

有什么影响：数据架构，互操作，数据安全等

Then there are significant cost savings. According to [Amazon](#), running a data warehouse on yourself costs between \$19k – 25k per terabyte per year. On average, companies see a 96% savings switching to a modern CDW (Cloud Data Warehousing).



# Data Fabric

---

混合云是市场的主流。众所周知，世界正迈入混合基础架构的云时代，我们需要管理各种类型的数据。利用Data Fabric在内部解决方案和云之间实现集成是取得成功的关键所在。

2014年NetApp最早提出这个概念。一直认为混合云是未来的首选，实际上从中国市场来看，企业主流毫无疑问也是公有云和私有云混合的。

NetApp的Data Fabric数据架构理念，就是通过搭建一个数据的管理架构，使这个数据可以很方便的在各个云之间流动，把私有云、公有云能够无缝联接起来。

有什么影响：数据架构，互操作，数据安全等

华为正式开源数据虚拟化引擎openLooKeng

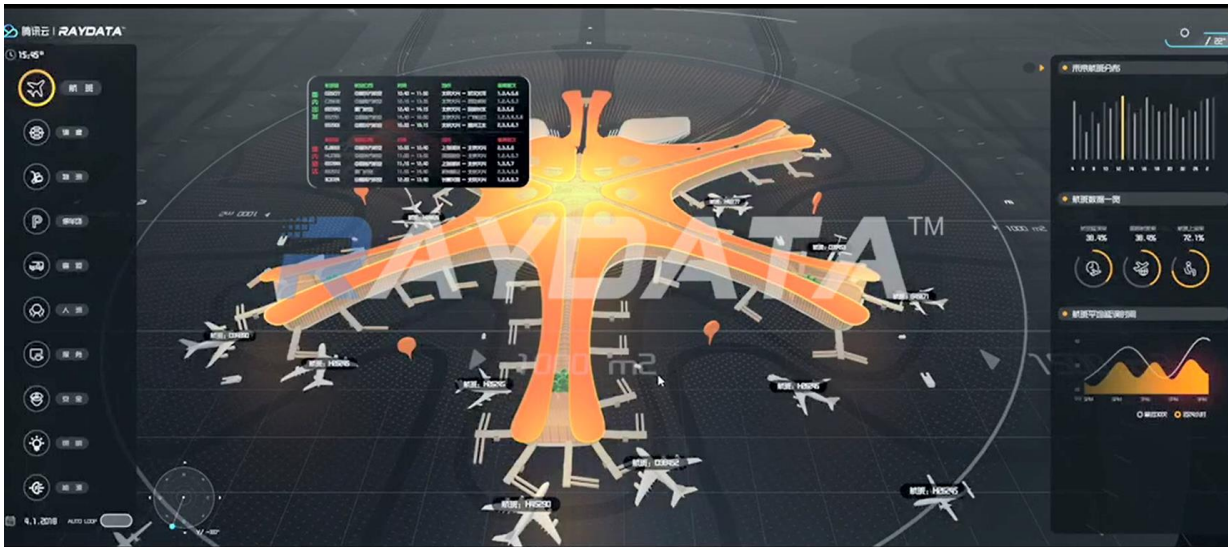
# 边缘计算：Edge Computing

---

虽然边缘计算历史上一直与物联网系统联系在一起，但Gartner相信，在整个2020年，该技术将成为所有行业和用例的“主导因素”。

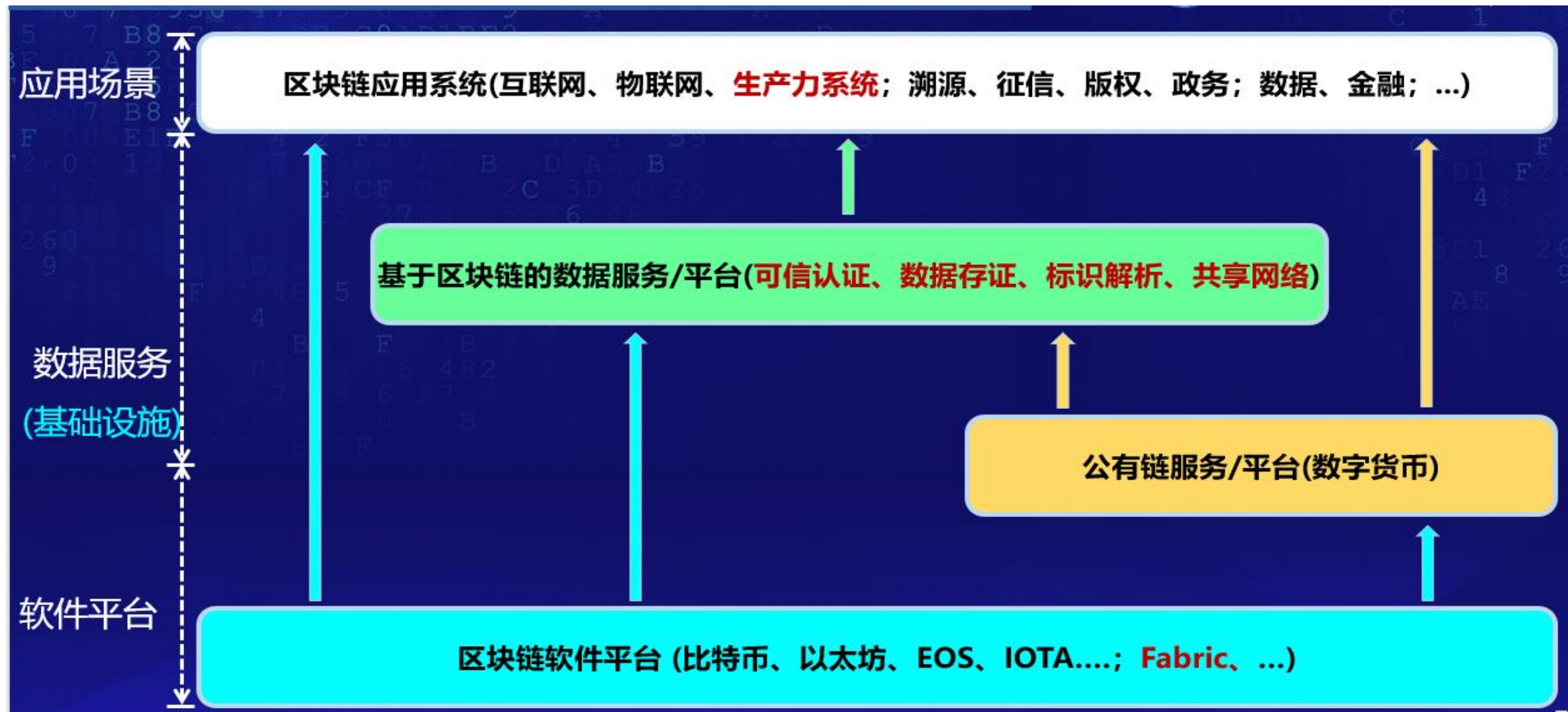
- 可以实时或者更快的进行数据处理和分析。让数据处理更靠近源，而不是外部数据中心或者云，可以缩短延迟时间。
- 大大减轻经费预算。企业在本地设备上的数据管理解决方案所花费的成本大大低于云和数据中心网络
- 减少网络流量。从而降低传输成本，破解数据传输瓶颈

有什么影响：数据架构，互操作，数据安全等



# 区块链的管理

Gartner认为区块链在降低成本、减少交易结算时间、改善现金流、提高透明度和可追溯性的能力方面有着巨大的潜力，同时还能产生更大的信任。



# 实时决策：RTD (Real Time Decision)

---

业务对实时性的需求越来越多。现在，由于云计算、流媒体软件的进步以及物联网（IoT）中传感器的增长，最终在更广泛的范围内实现这些系统（Gartner称之为持续智能 - Continuous Intelligence）是可行的。到2022年，超过一半的大型新业务系统将整合使用实时上下文数据改进决策的持续智能。

有什么影响：数据架构，流数据、敏捷等



# 新型数据构架

## 大数据中心核心构架



## 数据类别

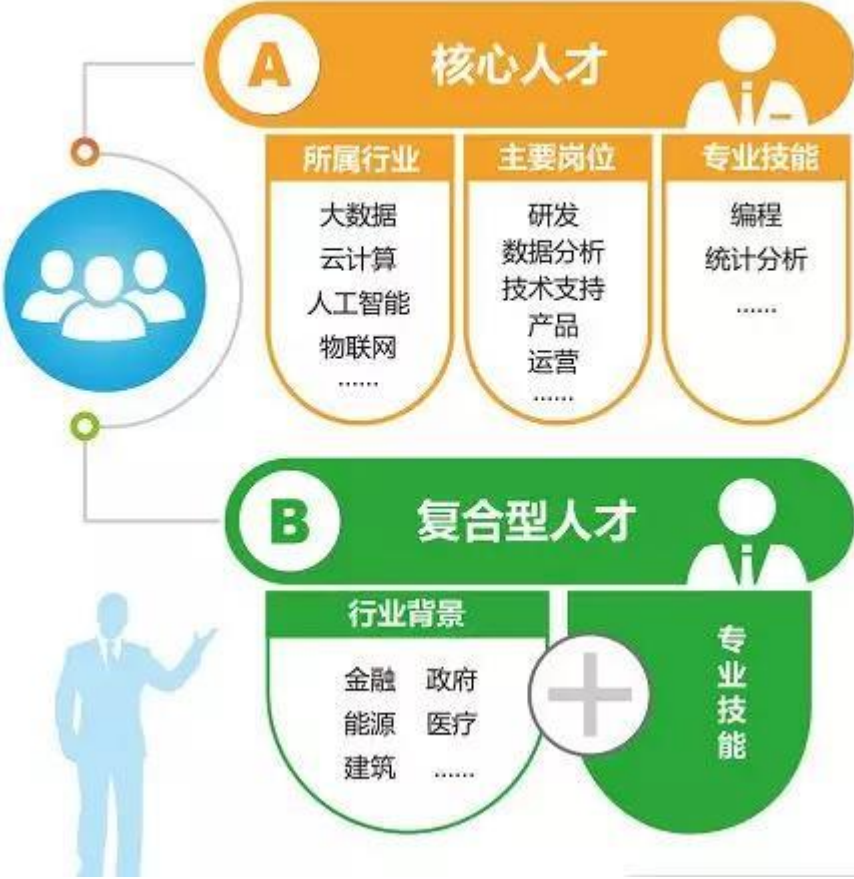


## 服务对象



# 新型混合人才和新型工种

赛迪智库 | 2025年中国大数据核心人才缺口将达230万



数据来源: [https://www.thepaper.cn/newsDetail\\_forward\\_4520721](https://www.thepaper.cn/newsDetail_forward_4520721)

# DAMA与国外教育行业合作

## □ 美国:

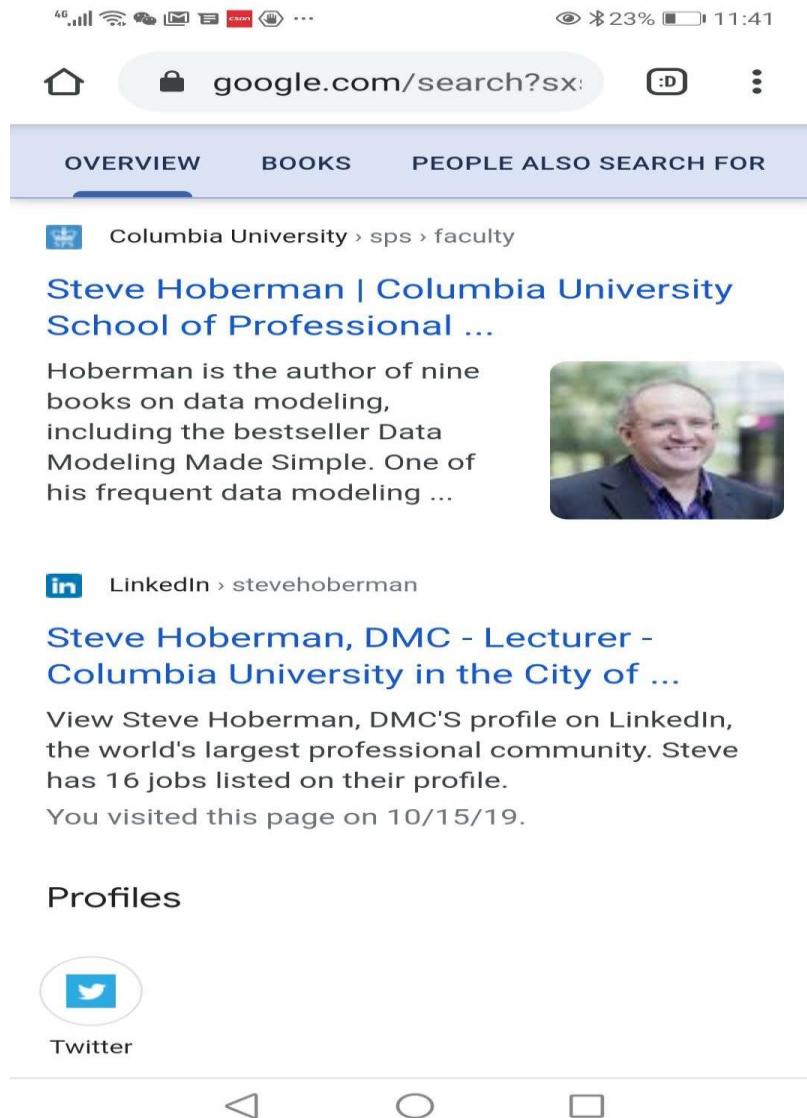
- ✓ 和哥伦比亚大学的合作
- ✓ 和UC Berkeley, MIT, Stanford, CMU的合作

## □ 英国

- ✓ 和爱丁堡大学的合作

## □ 日本

- ✓ 东京大学



# 趋势总结

---

- 新的分析和决策更需要数据管理。数据管理进入大爆发时期。
  - 实时性日趋重要，RTD的概念
  - 数据开放和共享，跨行业、跨领域、跨区域的数据平台
  - 通信和数据标准的建设
  - 边缘计算的广泛应用
  - 作为数据管理的一种特殊形态，区块链的应用一直在尝试
- 自主的ToB软件系统将蓬勃发展，特别是数据资产目录（Data Catalog)和数据资产价值评估
- 数据标准的快速发展和制定
- 数据安全会越来越受到重视，从GDPR到美国加州数据保护，从边缘计算到非结构化数据
- 技术的创新，人脑和认知科学将是非常重要的一个理论研究领域
- 跨界人才的培养，比如“数据机械工程师”



# 谢谢!

## 网址:

DAMA International: <http://www.dama.org>

DAMA China: <http://www.dama.org.cn>



## 电子邮件:

DAMA International: [info@dama.org](mailto:info@dama.org)

DAMA China: [info@dama.org.cn](mailto:info@dama.org.cn)



汪广盛, 15158055698 (微信同号)