

数据 服务框架

Data Service Framework

COSR 编写组◎著



数据 服务框架

Data Service Framework
COSR 编写组◎著

图书在版编目 (CIP) 数据

数据服务框架 / 吴桐等著. —北京: 中信出版社,
2016.3
ISBN 978-7-5086-5658-8

I. ①数… II. ①吴… III. ①金融-数据收集 ②金融-数据管理 IV. ①F830.41

中国版本图书馆CIP数据核字 (2015) 第 268191 号

数据服务框架

著 者: 吴 桐 等

策划推广: 中信出版社 (China CITIC Press)

出版发行: 中信出版集团股份有限公司

(北京市朝阳区惠新东街甲 4 号富盛大厦 2 座 邮编 100029)

(CITIC Publishing Group)

承 印 者: 北京画中画印刷有限公司

开 本: 880mm×1230mm 1/32

印 张: 6.25 字 数: 100 千字

版 次: 2016 年 3 月第 1 版

印 次: 2016 年 3 月第 1 次印刷

广告经营许可证: 京朝工商广字第 8087 号

书 号: ISBN 978-7-5086-5658-8 / F · 3531

定 价: 30.00 元

版权所有·侵权必究

凡购本社图书, 如有缺页、倒页、脱页, 由发行公司负责退换。

服务热线: 010-84849555 服务传真: 010-84849000

投稿邮箱: author@citicpub.com

● ● ● ● ● 作者简介

策略研究理事会 (Council of Strategy Research), 是一家致力于大数据行业发展的非营利性组织, 通过规范数据服务, 提升服务能力, 来优化决策的效率。由多家公司共同发起。

● 闻 中

清华大学经济管理学院管理科学与工程系助理教授、纽约大学信息系统博士、信息系统协会 (AIS) 会员和运筹学与管理学研究会 (INFORMS) 会员, 主要研究领域为电子市场、电子商务策略、信息产品定价、信息经济学。发表的期刊论文有 "Consumers' Search for Information on the Internet: How and Why China Differs from Western Europe", "Finitely loyal customers, switchers, and equilibrium price promotion"。工作论文有《广告投放规模、消费者流量与价格竞争》《B2B市场的形成和激励》《电子商务中的价格公平》《互联网上的竞争产品定制: 离散选择模型》《关于传统零售商决策网络化分析》等。

● 吴 桐

成都数联铭品科技有限公司首席财务官 (CFO), 资深财务经营者, 曾任兴蓉投资股份有限公司资深财务经理、华西证券场外市场部总监, 成功参与过公司战略融资和上市方案设计和实施; 专注于企业大数据服务领域, 建立大数据在审计、金融领域的应用模型。毕业于上海财经大学, 第一位在校本科期间注册会计师 (CPA) 全科通过的优秀毕业生; 国内最大知识分享社区网站 "知乎" 财经类第一达人, 拥有粉丝数超过 18 万, 并著有财经类畅销书《财务笔记》。

——王新锐

清华大学法学学士、北京安理律师事务所高级合伙人、北京市朝阳区律师协会青年联合会委员，主要业务领域包括风险投资、私募股权、公司并购和知识产权。关于IT行业的多篇文章发表于腾讯科技、财新网、钛媒体和《中国风险投资》杂志，2010年以后将业务重心转向媒体、科技和通信（TMT）行业，具体涉及电子商务、在线教育、问答社区、新媒体营销、网络游戏等多个细分领域，曾协助多家创业公司获得总金额达数千万美元的风险投资，并担任常年法律顾问。

——曾 途

成都数联铭品科技有限公司首席执行官（CEO），专注于通过大数据技术为企业提供数据决策支撑服务，具有资深的商业服务经验，曾为美国通用电气（GE）、日本电气公司（NEC）、美国国际商用机器公司（IBM）、爱立信、中国移动、中国联通、中国电信、华为和中兴等企业提供商业服务。

——易欢欢

宏源证券研究所执行所长、互联网金融千人会联合创始人兼秘书长、中国互联网金融博物馆理事、国际金融论坛互联网金融中心执行主任、中国财经青年学会执行主席。中国最具影响力的科技研究者实践者之一，直接引导了中国资本市场互联网金融、大数据、云计算三次大潮；撰写出行业第一本大数据专著《大数据时代的历史机遇》，主编了第一本《互联网金融》杂志，2013年撰写的《互联网金融——金融业门口的野蛮人》一文被传播超过800万次，成为影响社会发展的年度文章。

——黄 震

中央财经大学法学院教授、金融法研究所所长、《互联网金融》杂志总编辑、互联网金融千人会联合创始人、中国支付清算协会互联网金融专业委员会专家。多年从事金融法、互联网金融等领域研究，一直致力于推动中国互联网金融领域创新研究与法制建设，为各级政府和金融机构提供了非常多有价值的规划。2013年创立互联网金融千人会，创刊《互联网金融》杂志，发起互联网金融全球峰会；2014年出席诺贝尔奖经济学家中国峰会、全国政协互联网金融论坛、美国纽约中国论坛等重要会议，有力促进了中国互联网金融领域的创新繁荣和健康发展。

肖 冰

达晨创投合伙人兼总裁、2014 年位列年度中国最佳创业投资人 28 名。他出色的投资能力和卓著的投资业绩在业内享有盛名。经过 20 多年在创投行业内的历练，投资布局也愈加成熟；主导了蓝色光标、同洲电子、拓维信息、中汇医药、和而泰、亿纬锂能、网宿科技、爱尔眼科、拓尔思、恒泰艾普、数码视讯、雷杜生命、英威诺、爱尚鲜花、凯妙服饰、东信时代的投资。

徐宗本

中国科学院院士，西安交通大学数学与统计学院教授、博士生导师、国务院学位委员会数学学科评议组成员、国家科技部《国家重点基础研究发展规划》(973) 信息领域第二届咨询专家组成员、国家自然科学基金委员会“可信软件基础研究”指导专家组成员；主持完成国家 973、863、国家自然科学基金、香港用户生成内容 (UGC) 科学基金等国家级科研课题课 15 项。长期从事巴拿赫 (Banach) 空间几何理论与智能信息处理的数学基础方面的教学与研究，发表 SCI 学术论文 137 篇，将数据建模从“辅助作图”层次提升到“认知模拟”层次，系统提出了基于视觉认知的数据建模新原理与新方法；发现非欧氏框架下的“类二项式公式”数据建模新工具 (Xu-Roach 定理)，奠定了机器学习正则化方法的分析基础；提出压缩感知 $L_{1/2}$ 正则化理论，解决了稀疏信号处理、神经网络系统、模拟进化计算中的一些重要基础问题。

燃灯者

每一个试图阻挡新时代到来的人，都会成为旧时代的关灯者和关门人。

大数据落地的这两三年，一个崭新的时代已经到来。我们对以前一些传统行业的传统投资方式，传统的商业模式、传统的产品设想，都要能够去打破。

用数据来驱动科技的创新，带来翻天覆地的全新的商业价值，这是一种不可抵挡的趋势，可能来得比想象中还要快，但凡希望用传统方式，来阻挡甚至阻止这种趋势到来，都不会成功。

所以，我们要用更开放的心态来拥抱和迎接这样一个新时代的到来。

大数据时代有三大趋势：第一个是数据总量呈爆炸性增长。现在每天产生的数据量相当于公元 0 年大

约要一千年左右全人类产生的数据量总和，而且这个量还在不停地呈指数式增长。

海量的信息也让人难以选择。所以大数据时代的第一个重大矛盾：可以获取的数据总量爆炸性增长和普通人分辨甄别数据能力之间的矛盾，我们形象地称其为数据过载（**Information Overload**）。这个大问题，实际上也是很多大数据产业或者大数据创新可能有机会的地方。

第二个趋势：数据的形态发生了巨大变化，以前我们接触到的绝大多数数据叫作结构化数据，非常清楚，定义也很完整。面对这样的数据，运用简单的机器学习的办法就能得到各个数据之间的关联，从而做一些简单的预测。但是从去年到现在，超过85%的新数据都是非结构化的数据，蕴藏着巨大价值。它们和结构化的数据不一样，我们没有一种普适的方法去挖掘其价值，这就带来了第二个挑战：怎样挖掘非结构化数据之间的价值，甚至把它们转化为结构化的数据？

第三个趋势：数据的关联形态发生了变化。以前我们也有很多数据，线上线下的海量数据在不同的部门价值都非常大，但实际上，数据与数据之间相当于信息的孤岛，他们互相之间并不沟通。

时代的大潮不容逆转，乘风破浪，摧毁重建的过程，必然伴随着阵痛。但是，为了实现“拥抱”的自由，真正需要的是深刻的自律。对个人是这样，对于这个蓬勃发展的行业更是这样。

伊曼纽尔·康德对自由与自律曾做过经典的表述：“他认为人作为理性存在物，其意志应当是自由的，自由是人的一种天赋

权利，是每个人由于他的人性而具有的独一无二的、原生的、与生俱来的权利。一个人只有作为纯粹的意志存在物，而不是作为一个自然的存在物，他才是自我决定的，才能‘自己为自己立法’，才能在立法时不服从异己意志。”

标准化是大数据行业良性发展的重要基石，在大数据的国际化进程中，全球范围内大数据的标准化工作还处于研究阶段，无论是术语、架构，还是存储、处理、行业应用等，都尚未形成一种标准规范，因此，加快标准化研究和制定，规范大数据行业，推进行业发展，为我国的大数据战略层设计做参考，将我国的大数据诉求积极输出到国际标准组织中，是我国引领国际大数据标准化的关键一步。

COSR的产生，正是源于“自己为自己立法”，以期实现真正的行业规范和自律，从而推动整个大数据行业可持续健康发展。《战国策》里的名篇《触龙说赵太后》有云：父母之爱子，则为之计深远。当前，大数据行业还处在快速成长的阶段，作为关切行业发展的从业者，我深知，对这个行业最深沉的爱体现在推动它的合规合法，增强它的自身活力，走得更稳健，更长远。

“是知灯者，破愚暗以明斯道。”

诚愿COSR能成为新时代的燃灯者，为社会发展与文明传承保驾护航，让智慧薪火相传。

周涛

电子科技大学大数据研究中心主任

数据服务时代

这是一个被大数据彻底颠覆的时代！

我们已经迈入互联网时代：科技高速发达，信息持续积累，人与人之间的交流越来越密切，生活也越来越便利。在这便利的背后是人们的生活、工作、日常交际都以数字化形式呈现和记录，从而形成了大数据——这个伴随着时代发展而产生的高科技产物。

大数据到底有多大？一组名为“互联网上一天”的数据告诉我们：“一天之中，互联网产生的全部内容可以刻满 1.68 亿张 DVD；发出的邮件有 2940 亿封之多（相当于美国两年的纸质信件数量）；发出的社区帖子达 200 万个（相当于《时代》杂志 770 年的文字量）；卖出的手机为 37.8 万台，高于全球每天出生的婴儿数量 37.1 万……”怎么说“大”也不为过。

面对大数据，如果我们有相应的科技技术、分析手段驾驭它，那大数据就是无穷无尽的金矿；如果没有，那大数据就很有可能成为淹没我们的海洋。COSR（策略研究理事会）是一个致力于大数据行业发展的非营利性组织。该组织致力于通过规范数据服务、提升服务能力，来优化决策的效力和效率。COSR这一路走来已经可以客观审视对待已有的大数据分析基础，提前思考、完善构架、从容部署。

COSR提出的数据服务框架，旨在打造数据服务需求方、服务供给方和其他利益相关方共同认可、遵循的业务规范，实现数据服务的可靠性以及合规性，提升大数据服务市场的透明度，促进大数据服务市场的健康和迅速发展。从数据的采集、清洗、储存、分析、应用以及监督管理的全企业层面的BI（商业智能）平台，并培养贯穿企业运营管理流程的BA（商业分析）体系，力求以数据说话，实现全企业层面的高效管理和精准营销、销售、服务，也就是COSR可以最终成为大数据时代弄潮儿的关键所在。

这是一本关于大数据的及时雨的书。一本好书。

孙伟

北京软件行业协会执行会长

北京航空航天大学软件学院院长、教授

双龙航软创投基金主席

2015年12月14日于北京

DT 时代需要的变革与框架

马云说，人类正从 IT（Information Technology）时代，走向 DT（Data Technology）时代。所谓 DT 时代，就是大数据的时代。

“啤酒 + 尿布”的经典案例，让人们意识到数据背后潜藏的商业应用价值。而以这一案例为代表，Big Data（大数据）的出现，也颠覆了我们几千年来的一個基本概念：基于数据分析寻求因果关系，转向寻求关联性。

这是一个根本性的变化，一种价值观的颠覆。这种颠覆，必然会带来一场深远的变革。甚至，我们可以把这场变革和当年的发明蒸汽机带来工业文明来相提并论。

毫无疑问，大数据将是我国信息化发展步入深水

区后的核心主题和战略抉择。最近一段时间，不论是国务院出台《关于促进大数据发展的行动纲要》，还是十三五规划首提“实施大数据战略”，都在搭建着大数据的顶层设计。

我们相信，大数据的崛起，必将带来一场新的商业变革。

大数据的时代来临，除了人才以外，企业最值钱的便是数据资产。在BAT无所不能的背后，正是他们掌握的海量数据。

在早期投资中，达晨创投主要以新兴行业为主，比如互联网和医疗健康行业。而现在，随着移动互联网和物联网的发展，数据量越来越大，能做的事情越来越多。所以，沿着大数据产业来做布局，我们几乎是以买跑道的方式来做投资的。

我们希望，通过资本的力量，围绕着大数据产业的本质，构建达晨大数据的“一带一路”。“一带”，就是从数据源、数据存储，包括数据算法、数据安全等方面，构建大数据产业的底层框架；“一路”，就是围绕已经投资的近300家企业，拥抱大数据的应用构建。

当然，任何新兴行业的发展，都是不断地完善和提升的过程。经历了短暂的泡沫之后，大数据产业目前正处于快速发展的阶段。目前来看，贴着大数据标签的，炒作大数据概念的，还大有人在，对整个行业的发展也有着不利的影响。与此同时，面对高度复杂的大数据服务，大数据需求企业很难判断数据服务质量，无法有效规避潜在的经营和法律风险。

COSR数据服务框架的意义，恰恰就是给行业建立起一个大企业治理结构的框架标准。因为它满足了对数据的两个普适性的要求：可靠与合规。

就是说，当你需要数据服务的时候，它是可用的，其标准、质量，甚至其所需要的时间，是达到一定的可靠标准的。同时，你不用担心数据提供方会有法律问题，数据的采集、整理、存储、分析和运用都符合相关法律法规的要求，没有不可控的法律风险。

在阐述IT与DT的差别时，马云认为，IT时代以“我”为中心，DT时代则以“别人”为中心，让别人更强大，开放和承担更多的责任。不同于传统资产的独占性，数据资产的价值来自流动。如果数据无法流动，大数据的价值就无法得到彻底实现。COSR数据服务框架打通了数据流通的瓶颈，堪称商业大数据领域的ISO9000。

与此同时，这应该是一个持续改进的框架。伴随着大数据企业经营的提升，新的大数据商业模式的出现，需要将新的要素整合到这一框架中来。

大数据行业本身亦是如此，它需要在不断地自我否定和肯定中前进，不断地探索良性的商业发展模式。

大数据时代带来整个社会的大变革，资本市场不可避免地处于变革当中。当前，中国资本市场处于空前繁荣的时期。在繁荣的背后，也可能出现任何市场都难以避免的泡沫。我们希望这种泡沫能够成为创新的催化剂和推动力，当泡沫破灭的时候，有一大批新兴行业的企业、伟大的企业能够破土而出。这其中，不应缺少大数据企业。

肖冰

达晨创投合伙人兼总裁

● ● ● ● ● 前 言

大数据方兴未艾。信息革命走到今天，人类第一次可以高速处理海量信息，大数据成为各行业发掘新的业务洞察、支持决策优化的战略武器。据《2014—2018年中国大数据产业发展前景与投资战略规划分析报告》，93%的企业都认为数据具有或可能具有很高的价值，大数据的价值已经在企业中获得了广泛的认可。

然而，大数据带来的不仅仅是机遇，在技术和应用层面实现大数据的价值面临着很多挑战和困难，例如：数据来源良莠不齐、数据管理困难、数据分析模型构建不易、数据隐私和用户使用权难以平衡。为了解决这些问题，近年来出现了一大批致力于数据服务的专业数据服务企业，大数据服务市场应运而生。

作为新兴的高科技服务业务，大数据服务高度复杂。大数据需求企业很难判断数据服务质量，从而规

避潜在的经营和法律风险。在此背景下，COSR（策略研究理事会）提出了数据服务框架，旨在打造数据服务需求方、服务供给方和其他利益相关方共同认可、遵循的业务规范，实现数据服务的可靠性以及合规性，提升大数据服务市场的透明度，促进大数据服务市场的健康和迅速发展。

COSR全称为“Council of Strategy Research”，是一个致力于大数据行业发展的非营利性组织。该组织致力于通过规范数据服务、提升服务能力，来优化决策的效力和效率。COSR的发起者包括数联铭品科技有限公司、安理律师事务所、华院数据技术有限公司、国信优易数据有限公司、数据堂科技有限公司、中证信用增进股份有限公司、达晨创业投资有限公司、泰一指尚科技有限公司、美林数据技术股份有限公司、清华大学经济管理学院、北京航空航天大学软件学院、北航铭品大数据创新实验室、电子科技大学大数据研究中心、徐宗本院士、孙伟教授、周涛教授。

作者简介	VII
序言一	XI
序言二	XV
序言三	XVII
前 言	XXI
数据服务框架概要	001
数据服务框架内容	007
1 数据服务的定义	009
2 服务环境	011
框架与概念	011
关键要素	020
基本配置	020
3 基础数据管理	023
框架与概念	023
关键要素	027
基本配置	028