

油气大数据 分析利用

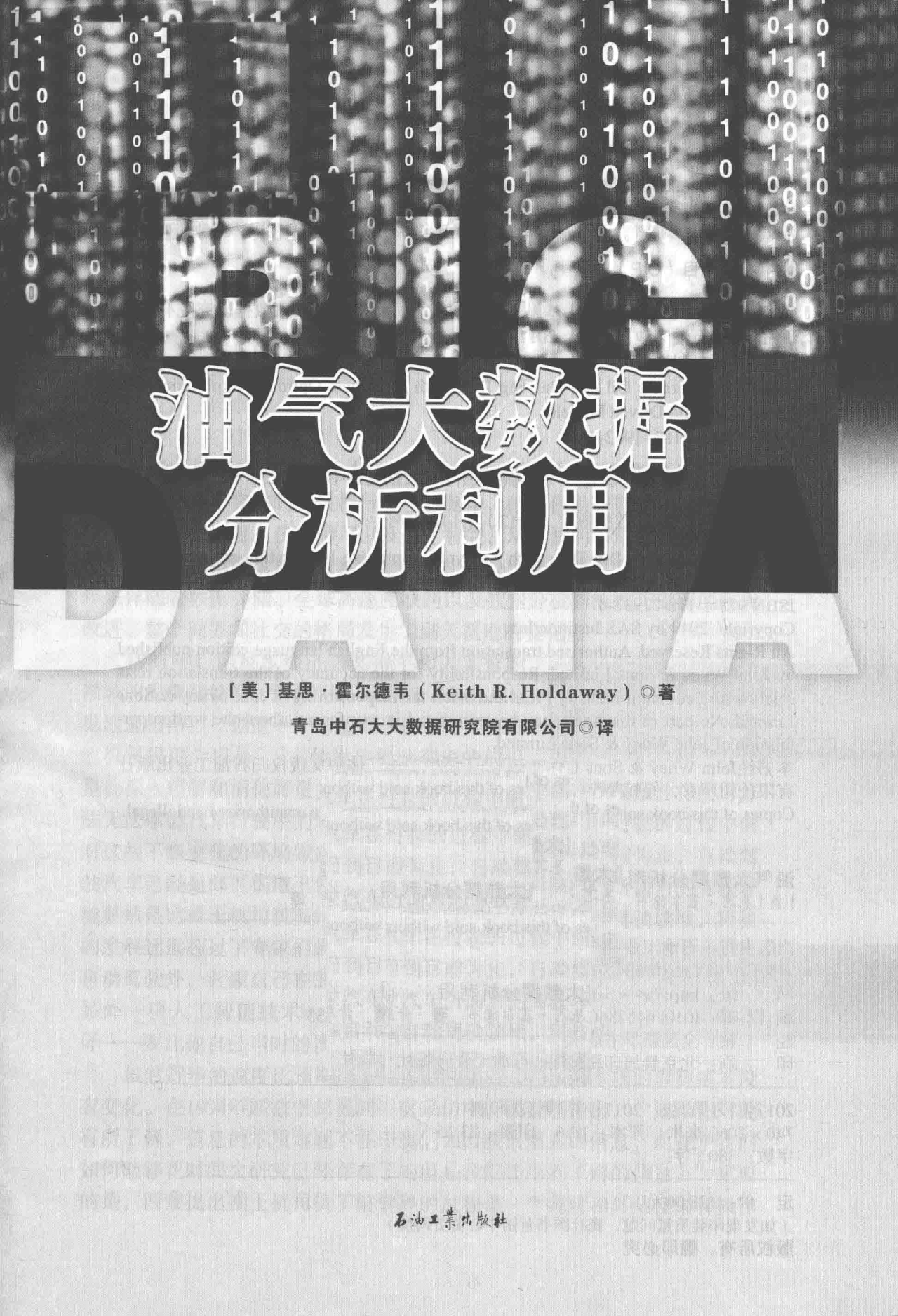
[美] 基思·霍尔德韦 (Keith R. Holdaway) ◎ 著

青岛中石大大数据研究院有限公司 ◎ 译

Harness Oil and Gas Big Data with Analytics
Optimize Exploration and Production
with Data-Driven Models

提供了大数据和分析技术应用于石油和天然气行业的完整视图
有效推动石油和天然气的勘探和生产

石油工业出版社



油气大数据 分析利用

【美】基思·霍尔德韦 (Keith R. Holdaway) ◎著

青岛中石大大数据研究院有限公司◎译

石油工业出版社

内 容 提 要

本书提供了大数据和分析技术应用于石油和天然气行业的完整视图。本书包含多个实例研究,强调了在石油和天然气勘探和生产阶段对优化的迫切需求,并展示了数据分析如何提供这种优化。本书涵盖了勘探和生产阶段石油和天然气行业面临的主要问题,揭示了如何建模大数据以实现效率和业务收益,涉及石油和天然气的勘探、开发、生产和油气资产的复兴,可有效地推动石油和天然气的勘探和生产。

图书在版编目(CIP)数据

油气大数据分析利用/[美]基思·霍尔德韦著;青岛中石大大数据研究院有限公司译.—北京:石油工业出版社,2017.7

(读点石油财经丛书)

书名原文:Harness Oil and Gas Big Data with Analytics: Optimize Exploration and Production with Data-Driven Models

ISBN 978-7-5183-1962-6

I. ①油… II. ①基… ②青… III. ①油气开采-数据处理 IV. ①TE3

中国版本图书馆CIP数据核字(2017)第141145号

Harness Oil and Gas Big Data with Analytics: Optimize Exploration and Production with Data-Driven Models by Keith R. Holdaway

ISBN 978-1-118-77931-6

Copyright 2014 by SAS Institute Inc.

All Rights Reserved. Authorised translation from the English language edition published by John Wiley & Sons Limited. Responsibility for the accuracy of the translation rests solely with Petroleum Industry Press and is not the responsibility of John Wiley & Sons Limited. No part of this book may be reproduced in any form without the written permission of John Wiley & Sons Limited.

本书经John Wiley & Sons Limited授权翻译出版,简体中文版权归石油工业出版社有限公司所有,侵权必究。

Copies of this book sold without a Wiley sticker on the cover are unauthorized and illegal.

油气大数据分析利用

[美]基思·霍尔德韦 著 青岛中石大大数据研究院有限公司 译

出版发行:石油工业出版社

(北京市朝阳区安华里二区1号楼100011)

网 址: <http://www.petropub.com>

编辑部: (010) 64523602 图书营销中心: (010) 64523633

经 销: 全国新华书店

印 刷: 北京晨旭印刷厂

2017年7月第1版 2017年7月第1次印刷

740×1060毫米 开本: 1/16 印张: 23.25

字数: 380千字

定 价: 128.00元

(如发现印装质量问题,我社图书营销中心负责调换)

版权所有,翻印必究

最近一段时间以来，技术进步最为突出的要说是大数据分析的爆炸性增长了。一个完整的大数据行业已经诞生，他们借助于广泛使用的互联网，利用数据挖掘、机器学习以及人工智能，从大量的可用且可操作的数据中提取有价值的信息。虽然数字计算机本身已经存续了几十年，但是近年来伴随着数据存储、全球高速互联网以及数据分析算法这几方面的巨大改进，整个商务和社交的格局发生了翻天覆地的变化。

1981年，诺贝尔经济学奖得主和人工智能领域的先驱赫伯特·西蒙（Herbert A. Simon）有见地地指出：“创造一个人造大学教授比人造推土机司机更为容易”^①。他发出如此评论的前提

① IEEE Computer, 1981年11月, p. 69-74.

是，深入理解和消化海量多变的数据是非常复杂的，这种复杂性将使得算法无法掌握汽车行驶中的驾驶环境，并且无法指挥汽车在行驶的过程中面对这些不断变化的环境做出连续的决策和判断。然而到目前为止，自动驾驶汽车已经是郊区街道上常见的存在，这些自动驾驶汽车所面对的驾驶环境显然是比推土机司机面对的驾驶环境更为复杂！在自动驾驶领域，科技的发展远远超过了专家们的预期。事实上，除了自动驾驶外，西蒙自己在20世纪90年代观察到的另外一项人工智能技术——自然语言的机器翻译——要比他自己当时的预期发展得更为迅猛^②。

② 参见1994年6月8日在匹兹堡邮报上的采访

虽然进步的速度比预期的要快，但对于数据的结构本质的理解基本没有变化。在1994年匹兹堡邮报同一次采访中，西蒙还指出，“如果你对此有所了解，信息的本质难题不在于我们如何获取更多的信息，而在于我们如何能够花时间去研究已经存在了的但是我们还不甚了解的信息。”重要的是，西蒙提出推土机司机了解世界的过程是一个眼睛和耳朵搜集和操作

复杂的数据，并将其简化到人脑可以实时管理的过程。将多数据流抽象成索引和描述符号是人类大脑处理信息的方式。现在最前沿的数据分析的算法和技术与上述人脑处理信息的方式高度类似。

在数据分析领域迅猛发展的同时，国际石油天然气行业也在几个方面进行平行扩张。这几个方面包括开发与部署同数据库结合的传感器和测量技术，以及数据存储检索工具。石油工业一直是提高能源回收效率新技术的早期采用者，现在对于数据分析这个工具也不例外。将数据转化为可以提高生产效率的信息（包括勘探和开采）对于企业来说是很有一价值的一件事。过去二三十年来，已经部署了包括永久井下测试仪、SCADA单元、海底完井、四维地震网络等几个主要数据采集系统。应用算法和技术来利用这些数据，并将其用于谋求企业乃至整个地球的福祉，是本书的重点。

基恩·霍尔德韦在上游石油行业和大数据分析方面都有直接的经验 and 造诣。作为在石油领域应用大数据分析技术的先驱者，他在石油和天然气勘探和开采的许多方面都曾开发过最先进的算法。作为地球物理学家，作者曾直接处理各种勘探数据，譬如地震信号（石油工业里最原始和经典的“大”数据之一），后来他的注意力转移到油藏工程、钻井以及生产等方面。本书着重于应用，包括全方位地采用一切必需手段把石油和天然气资源送到消费者手中的各种应用。这些应用包括勘探、油藏描述、钻井、油藏管理、产量预测，以及产量优化。为了不让读者对算法中各种数学公式产生困惑，作者解释了什么样的数据分析技术可以应用在什么样的特定问题上，并且加以油气行业的实例进行说明。

作为一个关于这个主题的很及时的内容，这本书的含金量很高。尤其是对于那些渴望应用新的数据分析的方法去帮助开发油气资源的人，这本书将是非常有帮助的。



2017年4月

（罗兰德·霍恩，Roland N. Horne，美国加利福尼亚州斯坦福大学托马斯·戴维斯·巴罗地球科学教授，普雷科特能源研究院资深研究员，石油工程师协会（SPE）荣誉会员奖获得者，美国国家工程院院士。）

非常荣幸能拜读基恩·霍尔德韦先生的这本书，受邀为此书中文版作序是我莫大的荣幸，但不免惶恐惴惴之至。近年来信息科技发展日新月异，令人目不暇接，大数据技术无疑是其中最炙热的一个。目前，中国从国家到地方政府相继出台了有关扶植政策，各行各业都在积极探索利用大数据技术服务于本行业，“大数据时代”“大数据是新的石油”的说法逐渐盛行。

对于投身石油工业的人来说，自然关心何为大数据、它与石油工业的关系，以及我国石油工业如何应用大数据技术并享受其带来的红利等？其实这些问题，也正是当下其他所有行业共同思考的问题。石油工业产业链条长，其勘探、开发、储运、炼化、销售等诸多板块不断产生并沉淀着海量数据，这些数据涉及学科众多、来源广泛，数据格式复杂，无论从大数据的“3V”“4V”还是“nV”等哪种定义，石油工业所富含的数据都应该是符合大数据描述特征的，即便从更小的链条尺度来观察也是如此，这就给“石油人”应用此技术带来了巨大想象空间。不过，迄今国内石油界整体上对大数据技术的开发利用还刚起步，远不及身边金融、电商等领域深入，从本行业的具体应用能得到的体验不多，多数人对大数据的理解还多来自于一些社会上经典案例的了解。大数据能真正像石油一样从用作灯油到逐步变成工业的血液，这个道路还在起步探索。

与此同时，国际油价持续低迷，并成为常态，如果能综合运用大数据、物联网和人工智能等技术，释放石油大数据中蕴含的潜在价值，为行业生产与经营决策提供更加全面和准确的依据，提高现有资源的调配效率，有效降低生产成本，这样的前景无疑是非常值得期待的。美国的沃森系统已可以根据医学资料和期刊信息来分析某种癌症并给出治疗方案，大大提高了服务质量，并能找出一些新的原来很难被发现的内在联系和规

律。不久的将来，我们石油工业中类似的应用也许会成为常规工具，它有助于我们更快、更富洞察力地在浩瀚的石油数据中挖掘出新的价值。

现在，国内外很多石油企业、科研院所、信息及咨询行业公司已开展了面向石油大数据技术的研究及应用工作。相比较而言，国外同行取得的成绩，值得我们虚心学习。本书作者基恩·霍尔德韦先生是石油大数据技术研究及应用的专家，此书也是国内目前已引进的为数不多的石油大数据领域专著，介绍了大数据技术在油田勘探开发等方面的诸多应用，内容丰富簇新，非常宝贵。感谢基恩·霍尔德韦先生的辛勤奉献！同时也要感谢我校大数据研究院能及时发现并翻译这本书以飨中国读者。



2017年4月

（查明，工学博士，教授，博士生导师，享受国务院政府特殊津贴。
2002年8月任中国石油大学（华东）党委常委、副校长。）

写这本书的动机来自于过去七年在上游油气工业，我所观察到的一些累积难题在当前已经普遍。其中三个主要的问题就是数据管理、量化地下不确定性和现场工程策略的风险评估。随着不同的工程领域海啸般数据的涌现，很明显，数据驱动模型能够提供惊人的洞察力，把原始的大数据变成可操作掌控的知识。我看到过当地球科学家不可避免地得出传统的定性和解释方法已无法作为单一方法来从勘探开发数据链的大数据中获取最大价值的结论后，他们转而采用整合软计算的分析方法。

一些关键技术论文集中详述了各类数据驱动模型和预测模型，一系列软计算方法也已显现出可信性，这使得大数据分析不再是随机猜测和不确定的。美国石油工程师协会发现，在全球性会议上涌现出的大量论文有力证实了神经网络、模糊理论和遗传算法成功应用在储层建模和模拟领域。随着老一辈地球科学家从石油工业的退出，新一代更青睐统计学和软计算方法的地球科学家涌现，我们将会在上游领域演化出更广泛的应用。当挖掘隐藏样式改进传统研究时，离散分布的智能井的数字油田产生了大量数据过剩。应用数据驱动建模的基本思想正在地球科学家和工程师间越来越流行。

当前油气行业正面临数据爆炸，这本书于此迫切时机应运而生。我们已发现储层表征中三维地震数据的叠前分析与多地震属性的提取结合不断增加。随着永久安装在洋底和非常规储层（页岩、煤层气、蒸汽辅助重力驱以及深海资产）钻井中的传感器的出现，我们正见证着数据密集型活动遍地开花的结果。

软计算概念包含一些启示性信息。这意味着什么呢？我们可采用混合分析流程来解决一些上游产业最具挑战性的难题。结合正淡出石油工业的专业知识以及用来探索和预测对资本支出和运营成本有负面影响的事件的

数据驱动模型，通过建立一个将软技能和专业知识与在任何油气作业中都最为关键的数据结合起来的卓越的协同分析中心来保留多年的实践经验。

在此，我想借此机会感谢此书稿的所有参与者和评阅者，尤其是Horia Orenstein在预测分析领域的精益求精的专业知识和Moray Laing极有价值的反馈、钻井的专业知识和对一些阐述案例研究图片的贡献。SAS研究所的Stacey Hamilton是位让人备受鼓舞的、耐心的编辑，对这本书的完稿功不可没。在此我还要感谢给我很多建设性意见的工业界同事，尤其是沙特阿美的Mike Pittman和SAS研究所的Mohammad Kurdi、David Dozoul和Sebastian Maurice，确保了其内容与工业的相关性和适用性。

第1章 软计算基础 / 1

- 上游数据分析中的现状 / 3
- 数据驱动模型 / 6
- 从柏拉图到亚里士多德的进化 / 11
- 描述性和预测性模型 / 12
- SEMMA 过程 / 15
- 高性能分析 / 16
- 上游数据的三个原则 / 20
- 探索与生产价值提案 / 22
- 油田分析 / 24
- 我是一名…… / 29
- 参考文献 / 32

第2章 数据管理 / 35

- 勘探与生产价值定位 / 38
- 数据管理平台 / 40
- 数据库阵列 / 49
- 结构化数据和非结构化数据 / 53
- 提取、转换和加载过程 / 54
- 大数据分析 / 56
- 标准数据来源 / 57
- 案例研究：生产数据质量控制框架 / 59

最佳实践 / 61
参考文献 / 66

第3章 地震属性分析 / 67

勘探与开发价值主张 / 69
时延地震勘探 / 70
地震属性 / 71
油藏表征 / 73
油藏管理 / 74
地震数据分析 / 75
案例研究：由地震属性定义的油藏特征 / 97
参考文献 / 113

第4章 油藏描述和模拟 / 115

勘探与生产价值主张 / 118
探索性数据分析 / 120
油藏描述周期 / 123
传统数据分析 / 124
油藏模拟 / 125
案例分析 / 131
参考文献 / 148

第 5 章 钻井和完井优化 / 149

- 勘探与生产价值主张 / 152
- 工作流程 1：减少非生产时间 / 154
- 工作流程 2：钻井参数优化 / 161
- 案例分析 / 164

第 6 章 油藏管理 / 181

- 勘探与生产的价值定位 / 184
- 未来数字油田 / 187
- 卓越分析中心 / 193
- 分析工作流程：最好的实践 / 196
- 案例分析 / 200

第 7 章 产量预测 / 223

- 勘探与开采的价值主张 / 226
- 基于网络的产量递减曲线分析工具 / 228
- 非常规油气储量估计 / 247
- 案例分析：加密井产量预测 / 250
- 参考文献 / 254

第 8 章 生产优化 / 255

勘探与生产价值主张 / 259

案例研究 / 259

参考文献 / 286

第 9 章 油气行业大数据分析与管理 / 289

勘探开发领域的价值主张 / 292

探索性数据分析模块 / 294

探索性数据分析方法中的统计图表与图形 / 300

细分集成方法 / 307

数据可视化 / 309

案例分析 / 312

第 10 章 大数据：结构化数据和非结构化数据 / 325

油气资源探测与生产 / 330

集合专家知识和数据驱动的系统 / 333

案例分析 / 338

多元地质统计学 / 347

大数据工作流程 / 349

参考文献 / 357

后记 / 358

第 1 章

软计算基础

天地之大，赫瑞修，比你所能梦想到的多出更多。

——莎士比亚《哈姆雷特》

石油和天然气工业已经见证了一个引人注目的争论，即过去十年中随着上游问题变得越来越复杂，以至于不能把确定性和解释分析方法委托于独立分离的学科。这个时候能否采用软计算技术？我们发现自己身陷横跨在勘探和生产价值链中间雪崩式的海量数据中，数据驱动模型开始从单纯的职业好奇心向油气行业自身的迫切需要转型。多学科分析方法的核心是数据挖掘，提供描述性和预测性模型以补充植根于第一原理的传统工程分析技术。数据的聚合、集成、不确定度量化和软计算方法等方面的进步，正在逐渐使得原本只是来源于各类独立分离的上游数据的补充信息能够更及时地创造出更准确的油藏模型。相比于基于数学、物理和专家经验的传统解释，软计算更易行、高效、稳健以及占用较少资源。

我们应该致力于探索丰富的软计算技术在石油工业中的应用所能给人类带来的多方面好处。

上游数据分析中的现状

什么是类人工智能？精确的定义很重要，但是许多专家明智地回答这个问题时说道，这个术语到目前为止还没有被精确定义。Bertrand Russell 曾经说过：

我不假装自己从精确的问题开始。我不认为有什么事情能从一开始

就非常精确。人们可以在探索的过程中逐渐达到精确¹。

从原始数据中获得的有关知识的断言，包括赋予精确的定义，始终得益于对一个特定的领域，如石油和天然气上游学科详尽的研究结果。我们正在关注到影响勘探和生产（E&P）价值链的四个主要趋势：大数据、云分析、社交媒体以及移动设备。这些驱动力量正在以不同的速度将地质学家引向实施应用软计算的技术。

E&P价值链上大数据的可视化需要使用美国数学家Tukey的全套的探索性数据分析表格、地图和图表²，来使多元和复杂的油气行业上游系统的数据中隐藏的模式和关系浮出水面。我们将在第3章，第4章和第9章详细地介绍这些可视化技术，因为这个技术对于油气行业中实施的数据驱动方法是至关重要的。

人工神经网络（ANN）、模糊逻辑（FL）和遗传算法（GA）是目前正在油气行业中包括油藏管理和模拟、生产和钻井优化、实时钻井自动化和设施维护等领域应用的类人工智能技术。支持数据驱动模型的数据挖掘技术在很多行业无所不在，并且在过去几年中，油气上游行业的工程师们根深蒂固和落伍的态度正在被现存的商业压力所稀释，这个商业压力即为探索和生产更多的油气能源来应对全球对能源不断增长的需求。

未来数字油田（Digital Oilfields of the Future，或DOFFs）和具有多重传感器及仪表的智能井正在高速产生大量的各种数据，这些数据非常有效地定义了像储层—生产井—进口设备集成系统这么复杂的和非均质的景观。这些高维度的数据被源自社交媒体活动的非结构化数据补充，在油气田现场操作和云计算中移动设备都被证明是有价值的，并呈现出高度的灵活性，在网络和数据中的表现不断得到提高。我们正处在一个理想的时代，可以将软计算方法与传统的确定性和解释性方法嫁接在一起。

大数据（Big Data）：定义

以下内容是应对一些横跨在E&P价值链上的大数据分析中遇到的固有挑战。根据定义，大数据是一个创造出来的科技新词，用于表示具有大容量的、复杂的、各种类型的以非常高的频率进行整合等特征的数据集合。大数据的这些特征导致了实质性的数据分析困难，而这些困难却不能由传统的数据处理方式和工具来解决。在关系数据库管理系统（DBMS）中使用大数据以及实现桌面统计和可视化软件上都有着明显的局限性。大