

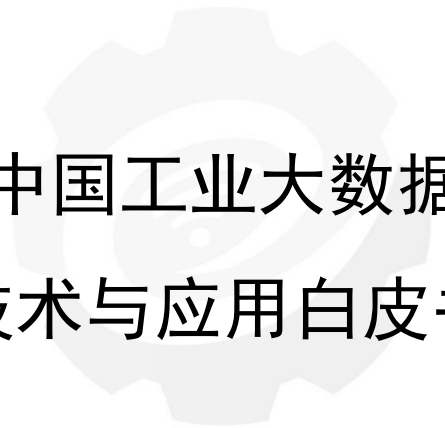


工业互联网产业联盟
Alliance of Industrial Internet

中国工业大数据 技术与应用白皮书

工业互联网产业联盟 (AII)

2017年7月



中国工业大数据
技术与应用白皮书

工业互联网产业联盟
Alliance of Industrial Internet

序言：拥抱工业大数据时代的到来

当前，以大数据、云计算、移动物联网等为代表的新一轮科技革命席卷全球，正在构筑信息互通、资源共享、能力协同、开放合作的制造业新体系，极大地扩展了制造业创新与发展空间。新一代信息通信技术的发展驱动制造业迈向转型升级的新阶段——数据驱动的新阶段，这是新的技术条件下制造业生产全流程、全产业链、产品全生命周期的数据可获取、可分析、可执行的必然结果，也是制造业隐性知识显性化不断取得突破的内在要求。

习近平总书记强调，“要着力推动互联网与实体经济深度融合发展，以信息流带动技术流、资金流、人才流、物资流，促进资源配置优化，促进全要素生产率提升”。习总书记这段话深刻阐释了互联网与实体经济的关系，阐释了以互联网为代表的新一代信息通信技术融合创新推动实体经济转型升级的内在机理，也充分体现了工业大数据作为一种新的资产、资源和生产要素，在制造业创新发展中的作用。可以从三方面来理解。

首先，资源优化是目标。新一代信息通信技术与制造业融合主要动力和核心目标就是不断优化制造资源的配置效率，就是要实现更好的质量、更低的成本、更快的交付、更多的满意度，就是要提高制造业全要素生产率。从企业竞争

的角度来看，企业是一种配置社会资源的组织，是通过对社会资本、人才、设备、土地、技术等资源进行组合配置来塑造企业竞争能力的组织，是一个通过产品和服务满足客户需求的组织，企业之间竞争的本质是资源配置效率的竞争，这是任何一个时代技术创新应用永恒追求的目标。

其次，数据流动是关键。新一代信息通信技术是如何优化制造资源配置效率？信息流是如何带动技术流、资金流、人才流、物流？关键是数据流动。从数据流动的视角来看，数字化解决了“有数据”的问题，网络化解决了“能流动”的问题，智能化要解决数据“自动流动”的问题，即能够把正确的数据在正确的时间以正确的方式传递给正确的人和机器，能够把海量的工业数据转化为信息，信息转化为知识，知识转化为科学决策，以应对和解决制造过程的复杂性和不确定性等问题，在这一过程中不断提高制造资源的配置效率。

第三，工业软件是核心。工业大数据的核心在于应用，在于优化资源配置效率，其关键在于，数据如何转化为信息，信息如何转化为知识，知识如何转化为决策，其背后都有赖于软件，软件是人类隐性知识显性化的载体，软件构建了一套数据如何流动的规则体系，正是这套规则体系确保了正确的数据能够在正确的时间以正确的方式传递给正确的人和机器。工业软件作为一种工具、要素和载体，为制造业建立了一套信息空间与物理空间的闭环赋能体系，实现了物质生产

运行规律的模型化、代码化、软件化，使制造过程在虚拟世界实现快速迭代和持续优化，并不断优化物质世界的运行。

《工业大数据技术与应用白皮书》的形成，是工业互联网产业联盟对新形势下工业大数据应用需求、技术演进和实施策略的阶段性提炼和总结，希望与业界分享，共同推动我国工业大数据产业进步、应用繁荣。

是为序。



安筱鹏

2017年6月27日

工业互联网产业联盟
Alliance of Industrial Internet

编写说明

工业大数据是工业领域相关数据集的总称，是工业互联网的核心，是工业智能化发展的基础原料。《中国制造 2025》规划中明确指出，工业大数据是我国制造业转型升级的重要战略资源。为此，在工业互联网产业联盟的指导下，总体组下辖的工业大数据特设组主持编写了这本《工业大数据技术与应用白皮书》。

以“做有用的白皮书”为根本出发点，本书努力回答，工业大数据是什么？从哪里来？到哪里去？如何管理？如何分析？如何应用？等基本问题。第 1 章重点阐述工业大数据的概念、价值和历史由来；第 2 章重点介绍工业大数据的典型应用场景；第 3、4 章分别从业务和技术两个角度探讨了工业大数据的实施路径和系统架构；第 5、6 章介绍了工业大数据管理与分析的关键技术；第 7 章介绍了我国工业大数据典型应用案例；第 8 章展望了我国工业大数据的未来愿景。

本书成稿过程中得到了全联盟成员的大力支持，特别是在征集工业大数据应用案例过程中，联盟相关企业结合自身业务痛点深入分析总结了工业大数据应用过程中的做法和经验，为本书观点的形成与落地提供了实践支撑。

本书由工业大数据特设组组长单位清华大学与副组长单位中国信息通信研究院牵头编写。特别感谢清华大学孙家

广院士、航天科工李伯虎院士、工信部信软司安筱鹏副司长、中国信息通信研究院余晓晖总工、中国宝武集团研究院郭朝晖首席给予的全面指导与帮助。同时国务院发展研究中心产业经济研究部王晓明主任、中电湘计海盾周翔总经理、东方电气集团张启德首席、北京兮易信息技术有限公司颜强等在本书成稿过程中也提出了许多建设性意见，在此一并致谢。

当前，工业大数据仍处在高速发展的历史阶段，其概念内涵、技术方法、价值创造模式还在不断创新演化之中，由于时间和水平所限，本书还存在着许多缺点和不足，期待联盟成员和广大读者给予批评指正，以便不断补充完善。

牵头编写单位：清华大学，中国信息通信研究院

参与编写单位：冶金自动化研究设计院、陕西鼓风机（集团）有限公司、中联重科股份有限公司、中国大唐集团公司、北京工业大数据创新中心有限公司、上海仪电显示材料有限公司、东方国信科技股份有限公司、北京北科亿力科技有限公司

编写组成员：

清华大学：王建民、王晨、刘英博、任艮全

中国信息通信研究院：余晓晖、李海花、李铮、魏凯

冶金自动化研究设计院：孙彦广、徐化岩

陕西鼓风机（集团）有限公司：范骁龙、郎博

中联重科股份有限公司：周志忠、傅军

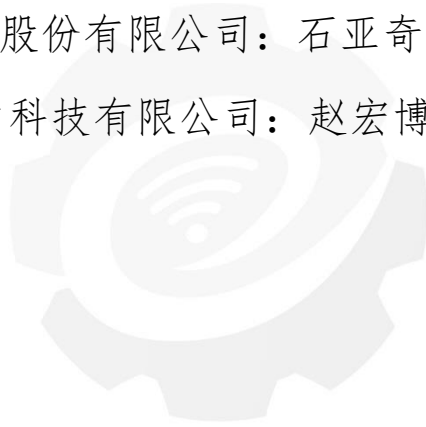
中国大唐集团公司：吕庭彦、丁罕

北京工业大数据创新中心有限公司：田春华、孙岚、钟
琥

上海仪电显示材料有限公司：蒋松涛、魏凤荣

东方国信科技股份有限公司：石亚奇、刘邦新

北京北科亿力科技有限公司：赵宏博、霍守锋



工业互联网产业联盟
Alliance of Industrial Internet

目 录

序言：拥抱工业大数据到来的新时代.....	2
1 工业大数据的概念与价值	1
1.1 工业大数据的内涵.....	1
1.1.1 工业大数据的定义.....	1
1.1.2 工业大数据的空间分布.....	2
1.1.3 工业大数据的产生主体.....	2
1.1.4 工业大数据的发展趋势.....	3
1.2 工业大数据的特点	4
1.2.1 工业系统的本质特征	4
1.2.2 工业大数据的 4V 特征.....	6
1.2.3 工业大数据的新特征	8
1.2.4 工业大数据应用特征	9
1.3 工业大数据的创新价值.....	12
1.3.1 数据始终影响着人类工业化进程	12
1.3.2 数据在信息化过程中发挥着核心作用.....	13
1.3.3 工业大数据是新工业革命的基础动力	13
1.4 工业大数据支撑中国制造弯道取直.....	14
1.4.1 中国是制造大国，但不是制造强国.....	14
1.4.2 工业大数据提升制造智能化水平，推动中国工业升级	15
1.4.3 工业大数据支撑工业互联网发展，促进中国工业转型	16
1.4.4 工业大数据助力中国制造弯道取直.....	17
2 工业大数据典型应用场景	18
2.1 优化现有业务，实现提质增效.....	18
2.1.1 研发能力提升	18
2.1.2 生产过程优化.....	18
2.1.3 服务快速反应.....	19
2.1.4 推动精准营销.....	19
2.2 促进企业升级转型.....	19

2.2.1	创新研发设计模式，实现个性化定制	20
2.2.2	建立先进生产体系，支撑智能化生产	20
2.2.3	基于全产业链大数据，实现网络化协同	22
2.2.4	监控产品运行状态和环境，实现服务化延伸	23
2.3	促进中小企业创新创业	24
3	工业大数据的实施策略	25
3.1	业务的加减乘除	25
3.2	过程的智能再造	26
3.3	架构的统筹规划	26
3.4	分析算法与模型的领域化	27
3.5	人才的培养引进	28
4	工业大数据系统与技术架构	30
4.1	工业大数据生命周期	30
4.2	工业大数据技术架构	32
4.3	工业大数据管理技术架构	33
4.4	工业大数据分析技术架构	34
5	工业大数据管理技术	36
5.1	多样性数据的采集技术	36
5.2	多模态数据的管理技术	36
5.3	高通量数据的写入技术	37
5.4	强关联数据的集成技术	38
6	工业大数据分析技术	40
6.1	强机理业务的分析技术	40
6.2	低质量数据的处理技术	41
6.3	数据高效率处理技术	42
7	工业大数据应用案例	43
7.1	北京工业大数据创新中心推进风电装备数字化升级	43
7.2	北科亿力推动炼铁行业大数据应用	49
7.3	大唐集团工业大数据应用	56
7.4	东方国信大数据实现联合利华能效提升	62

7.5 沙钢能源管理大数据应用	67
7.6 陕鼓动力智能运维大数据应用	74
7.7 仪电显示工业大数据应用实践	80
7.8 中联重科工业大数据应用实践	86
8、结语	93



工业互联网产业联盟
Alliance of Industrial Internet

1 工业大数据的概念与价值

本章主要讨论工业大数据的概念、意义和发展历程。

1.1 工业大数据的内涵

本节主要讨论工业大数据的内涵，从空间分布、产生主体两个维度对工业大数据进行分类，并讨论数据产生主体和内容结构演化路径。

1.1.1 工业大数据的定义

工业大数据即工业数据的总和，我们把它分成三类，即企业信息化数据、工业物联网数据，以及外部跨界数据。其中，企业信息化和工业物联网中机器产生的海量时序数据是工业数据规模变大的主要来源。

工业大数据是智能制造与工业互联网的核心，其本质是通过促进数据的自动流动去解决控制和业务问题，减少决策过程所带来的不确定性，并尽量克服人工决策的缺点。

首先，企业信息系统存储了高价值密度的核心业务数据。上世纪六十年代以来信息技术加速应用于工业领域，形成了制造执行系统（MES）、企业资源规划（ERP）、产品生命周期管理（PLM）、供应链管理（SCM）和客户关系管理（CRM）等企业信息系统。这些系统中积累的产品研发数据、生产制造数据、供应链数据以及客户服务数据，

存在于企业或产业链内部，是工业领域传统数据资产。

其次，近年来物联网技术快速发展，工业物联网成为工业大数据新的、增长最快的来源之一，它能实时自动采集设备和装备运行状态数据，并对它们实施远程实时监控。

最后，互联网也促进了工业与经济社会各个领域的深度融合。人们开始关注气候变化、生态约束、政治事件、自然灾害、市场变化等因素对企业经营产生的影响。于是，外部跨界数据已成为工业大数据不可忽视的来源。

1.1.2 工业大数据的空间分布

工业大数据不仅存在于企业内部，还存在于产业链和跨产业链的经营主体中。企业内部数据，主要是指MES、ERP、PLM等自动化与信息化系统中产生的数据。产业链数据是企业供应链（SCM）和价值链（CRM）上的数据，主要是指企业产品供应链和价值链中来自于原材料、生产设备、供应商、用户和运维合作商的数据。跨产业链数据，指来自于企业产品生产和使用过程中相关的市场、地理、环境、法律和政府等外部跨界信息和数据。

1.1.3 工业大数据的产生主体

人和机器是产生工业大数据的主体。人产生的数据是指由人输入到计算机中的数据，例如设计数据、业务数据、产品评论、新闻事件、法律法规等。机器数据是指由传感器、仪器仪表和智能终端等采集的

数据。智能制造与工业互联网发展，应致力于推动数据的自动采集。

对特定企业而言，机器数据的产生主体可分为生产设备和工业产品两类。生产设备是指作为企业资产的生产工具，工业产品是企业交付给用户使用的物理载体。前一类数据主要服务于智能生产，为智能工厂生产调度、质量控制和绩效管理提供实时数据基础；后一类数据则侧重于智能服务，通过传感器感知产品运行状态信息，帮助用户降低装备维修成本、提高运行效率、提供安全保障。

随着互联网与工业的深度融合，机器数据的传输方式由局域网络走向广域网络，从管理企业内部的机器拓展到管理企业外部的机器，支撑人类和机器边界的重构、企业和社会边界的重构，释放工业互联网的价值。

1.1.4 工业大数据的发展趋势

从数据类型看，工业大数据可分为结构化数据、半结构化数据和非结构化数据。结构化数据即关系数据，存储在数据库里，可以用二维表结构来表达实体及其联系。不方便用二维表结构来表达的数据即称为非结构化数据，包括办公文档、文本、图片、各类报表、图像、音频、视频等。所谓半结构化数据，就是以 XML 数据为代表的自描述数据，它介于结构化数据和非结构化数据之间。

二十世纪六十年代，计算机在企业管理中得到应用，经历了层次、网状等模型后，统一为关系模型，形成了以结构化数据为基础的 ERP/MES 管理软件体系。七十年代，随着计算机图形学和辅助设计

技术的发展,CAD/CAE/CAM等工具软件生成了三维模型、工程仿真、加工代码等复杂结构文件,形成了以非结构化数据为基础的PDM技术软件体系。二十一世纪,互联网和物联网为企业大量的文本、图像、视音频、时序、空间等非结构化数据,进而引发工业数据中结构化数据与非结构化数据的规模比例发生了质的变化。

近年来,智能制造和工业互联网推动了以“个性化定制、网络化协同、智能化生产和服务化延伸”为代表的新兴制造模式的发展,未来由人产生的数据规模的比重将逐步降低,机器数据所占据的比重将越来越大。2012年美国通用电气公司提出的工业大数据(狭义的),主要指工业产品使用过程中由传感器采集的以时空序列为主要类型的机器数据,包括装备状态参数、工况负载和作业环境等信息。

1.2 工业大数据的特点

在未来理想状态下,工业大数据应该作为工业系统相关要素在赛博空间的数字化映像、运行轨迹及历史痕迹。工业大数据的特点,应该体现工业系统的本质特征和运行规律,并推动工业进入智能制造时代。本节主要比较分析工业大数据的特点,进而为后续的讨论奠定基础。

1.2.1 工业系统的本质特征

工业系统往往具有复杂动态系统特性。飞机、高铁、汽车、船舶、火箭等高端工业产品本身就是复杂系统;产品设计过程,首先要满足

外部系统复杂多变的需求；生产过程更是一个人机料法环协同交互的多尺度动态系统；使用过程本质上就是产品与外部环境系统的相互作用过程。由此可见，产品全生命周期相关各个环节都具有典型的系统性特征。

确定性是工业系统本身能够有效运行的基础。对设计过程来说，确定性体现为对用户需求、制造能力的准确把握；对生产过程来说，确定性体现为生产过程稳定、供应链可靠、高效率和低次品率；对使用过程来说，确定性体现为产品持久耐用、质量稳定和对外部环境变化的适应性。因此，人们总是倾向于提高系统的确定性，避免不确定性因素对系统运行的干扰。工业系统设计一般基于科学的原理和行之有效的经验，输入输出之间的关系体现为强确定性。有效应对不确定性是工业系统相关各方追求的目标。工业系统是一个开放的动态系统，要面临复杂多变的内外部环境。因此，不确定性是工业系统必须面临的客观存在。工业产品全生命周期的各个阶段都面临着不确定性，例如外部市场与用户需求等因素的不确定性、制造过程中人机料法环等要素的不确定性，以及产品使用和运行环境的不确定性。

应对不确定性的前提是感知信息、消除不确定性。以工业互联网技术为代表的 ICT 技术的发展和普遍应用，能大大提升信息自动感知的能力，能让我们感知到用户需求和市场的变化、感知到远程设备和供应链的异动、感知到人机料法环等诸要素的状态，可减少人在信息感知环节的参与，降低人对信息感知所带来的不确定性影响。

在感知的基础上，可以更快速、科学地应对不确定性：通过智能

服务，解决用户使用过程中遇到的不确定性问题；通过智能设备，应对设备自身、原料以及运行环境所涉及的其它不确定性问题；通过智能生产，应对用户需求和工厂内部变化引起的不确定性问题；通过工业互联网，应对供应链、跨地域协同中的不确定性问题等等。在此基础上，相关过程产生的大数据，能够帮助我们更加深入、准确地理解工业过程，进而将工业过程中的个性化问题归结成共性问题、形成知识，并用于优化和指导企业的各种业务。这样，通过工业互联网和大数据技术的应用，能将不确定性转化为开拓市场、提质增效、转型创新的能力，把工业带入智能制造时代。

由此可见，工业系统同时具有确定性和不确定性的特征，确定性是目标，不确定性则是机会。

1.2.2 工业大数据的 4V 特征

工业大数据首先符合大数据的 4V 特征，即大规模（Volume）、速度快（Velocity）、类型杂（Variety）、低质量（Veracity）。

所谓“大规模”，就是指数据规模大，而且面临着大规模增长。我国大型的制造业企业，由人产生的数据规模一般在 TB 级或以下，但形成了高价值密度的核心业务数据。机器数据规模将可达 PB 级，是“大”数据的主要来源，但相对价值密度较低。随着智能制造和物联网技术的发展，产品制造阶段少人化、无人化程度越来越高，运维阶段产品运行状态监控度不断提升，未来人产生的数据规模的比重降低，机器产生的数据将出现指数级的增长。

所谓“速度快”，不仅是采集速度快，而且要求处理速度快。越来越多的工业信息化系统以外的机器数据被引入大数据系统，特别是针对传感器产生的海量时间序列数据，数据的写入速度达到了百万数据点/秒-千万数据点/秒。数据处理的速度体现在设备自动控制的实时性，更要体现在企业业务决策的实时性，也就是工业 4.0 所强调的基于“纵向、横向、端到端”信息集成的快速反应。

所谓“类型杂”，就是复杂性，主要是指各种类型的碎片化、多维度工程数据，包括设计制造阶段的概念设计、详细设计、制造工艺、包装运输等各类业务数据，以及服务保障阶段的运行状态、维修计划、服务评价等类型数据。甚至在同一环节，数据类型也是复杂多变的，例如在运载火箭研制阶段，将涉及气动力数据、气动力热数据、载荷与力学环境数据、弹道数据、控制数据、结构数据、总体实验数据等，其中包含结构化数据、非结构化文件、高维科学数据、实验过程的时间序列数据等多种数据类型。

所谓“低质量”，就是真实性（Veracity），相对于分析结果的高可靠性要求，工业大数据的真实性和质量比较低。工业应用中因为技术可行性、实施成本等原因，很多关键的量没有被测量、没有被充分测量或者没有被精确测量（数值精度），同时某些数据具有固有的不可预测性，例如人的操作失误、天气、经济因素等，这些情况导致往往数据质量不高，是数据分析和利用最大的障碍，对数据进行预处理以提高数据质量也常常是耗时最多的工作。

1.2.3 工业大数据的新特征

工业大数据作为对工业相关要素的数字化描述和在赛博空间的映像，除了具备大数据的4V特征，相对于其它类型大数据，工业大数据集还具有反映工业逻辑的新特征。这些特征可以归纳为多模态、强关联、高通量等特征。

“多模态”。工业大数据是工业系统在赛博空间的映像，必须反映工业系统的系统化特征，必须要反映工业系统的各方面要素。所以，数据记录必须追求完整，往往需要用超级复杂结构来反映系统要素，这就导致单体数据文件结构复杂。比如三维产品模型文件，不仅包含几何造型信息，而且包含尺寸、公差、定位、物性等其它信息；同时，飞机、风机、机车等复杂产品的数据又涉及机械、电磁、流体、声学、热学等多学科、多专业。因此，工业大数据的复杂性不仅仅是数据格式的差异性，而是数据内生结构所呈现出“多模态”特征。

“强关联”。工业数据之间的关联并不是数据字段的关联，其本质是物理对象之间和过程的语义关联。包括1) 产品部件之间的关联关系:零部件组成关系、零件借用、版本及其有效性关系；2) 生产过程的数据关联，诸如跨工序大量工艺参数关联关系、生产过程与产品质量的关系、运行环境与设备状态的关系等；3) 产品生命周期的设计、制造、服务等不同环节的数据之间的关联，例如仿真过程与产品实际工况之间的联系；4) 在产品生命周期的统一阶段所涉及到不同学科不同专业的数据关联，例如民用飞机预研过程中会涉及总体设计方案

数据、总体需求数据、气动设计及气动力学分析数据、声学模型数据及声学分析数据、飞机结构设计数据、零部件及组装体强度分析数据、系统及零部件可靠性分析数据等。数据之间的“强关联”反映的就是工业的系统性及其复杂动态关系。

“高通量”。嵌入了传感器的智能互联产品已成为工业互联网时代的重要标志，用机器产生的数据来代替人所产生的数据，实现实时的感知。从工业大数据的组成体量上来看，物联网数据已成为工业大数据的主体。以风机装备为例，根据 IEC61400-25 标准，持续运转风机的故障状态其数据采集频率为 50Hz，单台风机每秒产生 225K 字节传感器数据，按 2 万风机计算，如果全量采集每秒写入速率为 4.5GB/秒。具体来说，机器设备所产生的时序数据可以总结为以下几个特点：海量的设备与测点、数据采集频度高（产生速度快）、数据总吞吐量大、7X24 持续不断，呈现出“高通量”的特征。

1.2.4 工业大数据应用特征

我们把工业大数据的应用特征归纳为跨尺度、协同性、多因素、因果性、强机理等几个方面。这些特性都是工业对象本身的特性或需求所决定的。

“跨尺度”、“协同性”主要体现在大数据支撑工业企业的在线业务活动、推进业务智能化的过程中。

“跨尺度”是工业大数据的首要特征。这个特征是工业的复杂系统性所决定的。从业务需求上看，通过 ICT 技术的广泛深入应用，能将

设备、车间、工厂、供应链及社会环境等不同尺度的系统在赛博空间中联系在一起。事实上，工业 4.0 强调的横向、纵向、端到端集成，就是把不同空间尺度的信息集成到一起。另外，跨尺度不仅体现在空间尺度，还体现在时间尺度：业务上常常需要将毫秒级、分钟级、小时级等不同时间尺度的信息集成起来。为此，需要综合利用云计算、物联网、边缘计算等技术。

“协同性”是工业大数据的另外一个重要特征。工业系统强调系统的动态协同，工业大数据就要支持这个业务需求。我们进行信息集成的目的，是促成信息和数据的自动流动，加强信息感知能力、减少了决策者所面临的不确定性，进而提升决策的科学性。

“牵一发而动全身”是对“协同性”的形象描述、是“系统性”的典型特征。具体到工业企业，就是某台设备、某个部门、某个用户的局部问题，能够引发工艺流程、生产组织、销售服务、仓储运输的变化。这就要通过整个企业乃至供应链上多个部门和单位的大范围协同才能做到。

“多因素”、“因果性”、“强机理”体现在工业大数据支撑过程分析、对象建模、知识发现，并应用于业务持续改进的过程中。工业过程追求确定性、消除不确定性，数据分析过程就必须注重因果性、强调机理的作用。事实上，如果分析结果是具有科学依据的知识，本身就体现了因果性。

“多因素”是指影响某个业务目标的因素特别多。事实上，许多大数据分析的目标，就是去发现或澄清人们过去不清楚的影响因素。“多

因素”是工业对象的特性所决定的。当工业对象是复杂的动态系统时，人们必须完整、历史地认识考察它的全貌，才能得到正确的认识；对应到工业大数据分析，就体现为多个因素的复杂关系，进而导致了“多因素”的现象。

认清“多因素”特点对于工业数据收集有着重要的指导作用。人们往往需要事先尽量完整地收集与工业对象相关的各类数据，才有可能得到正确的分析结果、不被假象所误导。对于非线性、机理不清晰的工业系统，“多因素”会导致问题的维度上升、不确定性增加；对应到工业大数据分析过程中，人们常会感觉到数据不足、分析难度极大。

“因果性”源于工业系统对确定性的高度追求。为了把数据分析结果用于指导和优化工业过程，其本身就要高度的可靠性。否则，一个不可靠的结果，可能会引发系统巨大的损失。同时，由于工业过程本身的确性强，也为追求因果性奠定了基础。为此，工业大数据分析过程不能止步于发现简单的“相关性”，而是要通过各种可能的手段逼近“因果性”。

然而，如果用“系统”的观点看待工业过程，就会发现：系统中存在各种信息的前馈或者反馈路径。工业技术越是成熟，这种现象也就越普遍。这导致数据中体现的相关关系，往往并不是真正的因果关系。为了避免被假象迷惑，必须在数据准确完备的基础上，进行全面、科学、深入的分析。特别是对于动态的工业过程，数据的关联和对应关系必须准确、动态数据的时序关系不能错乱。

“强机理”是获得高可靠分析结果的保证。我们认为：分析结果的

可靠性体现在因果和可重复性。而关联关系复杂往往意味着干扰因素众多，也就不容易得到可重复的结论。所以，要得到可靠性的分析结果，需要排除来自各方面的干扰。排除干扰是需要“先验知识”的，而所谓的“先验知识”就是机理。在数据维度较高的前提下，人们往往没有足够的数据用于甄别现象的真假。“先验知识”能帮助人们排除那些似是而非的结论。这时，领域中的机理知识实质上就起到了数据降维的作用。从另外一个角度看：要得到“因果性”的结论，分析的结果必须能够被领域的机理所解释。事实上，由于人们对工业过程的研究往往相对透彻，多数现象确实能够找到机理上的解释。

1.3 工业大数据的创新价值

作为人类第一次自己创造的生产资料，工业数据一直伴随着工业的现代化进程，直至走入智能化阶段。

1.3.1 数据始终影响着人类工业化进程

恩格斯说：“任何一门学科的真正完善在于数学工具的广泛应用。”高质量、科学管理是工业企业走向现代化的前提。数据对提高质量、效率、管理的作用巨大。十八世纪末，画法几何学的创立标志着工程设计语言的诞生，伴随人类进入工业 1.0 时代，定量化、标准化成为工业 2.0 时代的主要特征，二十世纪中期，数字计算机在工业中的应用开启了工业 3.0 时代。从数据的发展历史看，数据由数、量演变而来，数据具有先天的精确性和实用性特征，计算方法与信息技术的应

用必然导致大数据的诞生。在信息化时代，数据随时随地与我们相伴而行，“数据密集型科学发现”已成为人类认知世界的第四范式。

1.3.2 数据在信息化过程中发挥着核心作用

随着工业进入信息化时代，数据成为工业系统运行的核心要素，追求的目标是把正确（R）的数据在正确（R）的时间，以正确（R）的形式送达给正确（R）的人。“三分技术、七分管理、十二分数据”、“垃圾进、垃圾出”等说法都说明了数据在信息化工程中的重要性。世界工业不断发展的过程，本质上是数据的作用逐渐加强的过程，数据在工业生产力不断提升的过程中发挥着核心作用。以自动化和信息化为代表的第三次工业革命以来，工业不断发展的过程也是数据传输和处理效率不断提高、数据质量不断提升、不确定性因素的应对效率不断加强的过程。通过建立包括产品定义数据、工艺数据、生产过程数据、在线监测数据、使用过程数据等在内的产品全生命周期数据治理体系，可以有效追溯质量问题的产生原因，并持续加强生产过程的质量保障能力。通过关联企业内外部多数据源的数据分析，可以挖掘发现复杂品质问题的根本原因。

1.3.3 工业大数据是新工业革命的基础动力

信息技术特别是互联网技术正在给传统工业发展方式带来颠覆性、革命性的影响。世界正加速进入一个互联互通的时代，互联网对工业的影响越来越深刻，并成为引发新一轮工业革命的导火索。互联

网技术全面深入发展，极大促进了人与人互联、机器和机器互联、人和机器互联的程度，随着 5G、量子通信等新一代通信技术发展，世界将加速进入一个完全互联互通的状态。工业互联网也将随着机器的数字化、工业网络泛在化、云计算能力的提高而取得长足进步，海量工业大数据的产生将是必然结果，而基于工业大数据的创新是新工业革命的主要推动力。

对于新工业革命而言，工业大数据就像是 21 世界的石油。美国通用电气公司的《工业互联网白皮书》中指出工业互联网实现的三大要素是智能联网的机器、人与机器协同工作及先进的数据分析能力。工业互联网的核心是通过智能联网的机器感知机器本身状况、周边环境以及用户操作行为，并通过这些数据的深入分析来提供诸如资产性能优化等制造服务。没有数据，新工业革命就是无源之水无本之木。工业互联网所形成的产业和应用生态，是新工业革命与工业智能化发展的关键综合信息基础设施。其本质是以机器、原材料、控制系统、信息系统、产品以及人之间的网络互联为基础，通过对工业数据的全面深度感知、实时传输交换、快速计算处理和高级建模分析，实现智能控制、运营优化和生产组织方式的变革。

1.4 工业大数据支撑中国制造弯道取直

1.4.1 中国是制造大国，但不是制造强国

中国制造业发展并不均衡，工业 1.0、2.0、3.0 并存，企业

较为普遍地存在着研发创新能力弱、产品质量差、能源消耗大、污染物排放多、管理粗放、自动化和信息化水平低等问题。中国制造业正在承受产业“双向转移”的压力。一方面，劳动密集型的以出口或代工为主的中小制造企业正在向越南、缅甸、印度和印尼等劳动力和资源等更低廉的新兴发展中国家转移。另一方面，部分高端制造业在美国、欧洲等发达国家“再工业化”战略的引导下回流。尽管中国制造业规模位居世界第一，但人均产值仅为美国的 16%，航空发动机、燃气轮机等高端装备制造受制于人的局面没有根本改变。中国制造业如不能快速转型升级，在高端产品制造尚未形成国际竞争力之前，中低端产品的制造竞争力也将被削弱，进一步面临“产业空心化”的风险。

1.4.2 工业大数据提升制造智能化水平，推动中国产业升级

大数据是提升产品质量、生产效率、降低能耗，转变高耗能、低效率、劳动密集、粗放型生产方式，提升制造智能化水平的必要手段。具有高度灵活性、高度自动化等特征的智能工厂是国际先进制造业的发展方向。广泛深入的数字化是智能工厂的基础。多维度的信息集成、CPS 的广泛应用与工业大数据发展相辅相成。通过推进智能制造，实现去低端产能、去冗余库存、降制造成本。结合数控机床、工业机器人等自动生产设备的使用，并建立从经营到生产系统贯通融合的数据流，做到数据全打通和数据流通不落地，可以提升企业整体生产效率，降低劳动力投入，有效管理并优化各种资源的流转与消耗。通过对设备和工厂进行智能化升级，加强对制造生产全过程的自动化控制和智

能化控制，促进信息共享、系统整合和业务协同，实现制造过程的科学决策，最大程度实现生产流程的自动化、个性化、柔性化和自我优化，实现提高精准制造、高端制造、敏捷制造的能力。大数据也是提升产品质量的有效手段。通过建立包括产品生产过程工艺数据、在线监测数据、使用过程数据等在内的产品全生命周期质量数据体系，可以有效追溯质量问题的产生原因，并持续改进生产过程的质量保障能力。通过关联企业内外部多数据源的大数据分析，可以挖掘发现复杂成因品质问题的根本原因。

1.4.3 工业大数据支撑工业互联网发展，促进中国工业转型

工业大数据是制造业实现从要素驱动向创新驱动转型的关键要素与重要手段。大数据可以帮助企业更全面、深入、及时了解市场、用户和竞争态势的变化，以推出更有竞争力的产品和服务。对于新产品研发，大数据不仅可以支持企业内部的有效协同、知识重用，还能利用众包众智等手段利用企业外部资源。这些做法，不仅能够提高研发质量、还能大大缩短研发周期。

大数据也是实现工业企业从制造向服务转型的关键支撑技术。通过产品的智能化，可以感知产品的工作状况、周边环境、用户操作的变化。在此基础上，可以提供在线健康检测、故障诊断预警等服务，以及支持在线租用、按使用付费等新的服务模型。通过对产品使用的实时工况数据、环境数据、过往故障数据、维修记录、零部件供应商数据进行整合，可以快速预判、实时掌握设备健康状况，减少设备停

机时间，削减现场服务人员；可以准确判断出现故障的潜在类型和原因，快速形成现场解决方案，缩短服务时间。

1.4.4 工业大数据助力中国制造弯道取直

《中国制造 2025》规划中明确指出，工业大数据是我国制造业转型升级的重要战略资源。有效利用工业大数据推动工业升级，需要针对我国工业自己的特点。一方面，我国是世界工厂，实体制造比重大；与此同时，技术含量低、研发能力弱、劳动密集、高资源消耗制造的问题相对突出，制造升级迫在眉睫。另一方面，我国互联网产业发展具有领先优势，过去十多年消费互联网的高速发展使互联网技术得到长足发展、全社会对互联网的重视度高。我们需要充分发挥这一优势并将之和制造业紧密结合，促进制造业升级和生产性服务业的发展。

目前，工业大数据已成为国际产业竞争和国家工业安全的基础要素相关技术与应用必将成为我国工业“由跟跑、并跑到领跑”、“弯道取直”、“跨越发展”的关键支撑。作为制造业大国，我国时刻产生着海量的工业数据。我们应该充分利用这一条件，创新管理思想、重构产业生态，提升中国制造在全球产业链分工中的地位。用工业大数据提高产品质量、管理水平、弥补在人员素质方面的差距，补齐落后的短板。在此基础上，推进智能制造和工业互联网的应用。利用我国工业门类齐全、互联网和电子商务应用的比较优势，力争在新工业革命时代实现换道超车。

2 工业大数据典型应用场景

工业大数据可应用于现有业务优化、推动大中型企业实现智能制造升级和工业互联网转型，并支撑中小企业创新创业。

2.1 优化现有业务，实现提质增效

工业大数据可以在现有组织、流程保持不变的前提下，把各个部门和岗位的工作做得更好，促进整个企业的提质增效。

2.1.1 研发能力提升

建立针对产品或工艺的数字化模型，用于产品、工艺的设计和优化。模型作为量化、可计算的知识载体，可提高企业知识重用水平、并促进持续优化。将大数据技术与数字化建模相结合，可以提供更好的设计工具，缩短产品交付周期。

例如，波音公司通过大数据技术优化设计模型，将机翼的风洞实验次数从 2005 年的 11 次缩减至 2014 年的 1 次；玛莎拉蒂通过数字化工具加速产品设计，开发效率提高 30%。

2.1.2 生产过程优化

通过分析产品质量、成本、能耗、效率、成材率等关键指标与工艺、设备参数之间的关系，优化产品设计和工艺。以实际的生产数据为基础，建立生产过程的仿真模型，优化生产流程。根据客户订单、

生产线、库存、设备等数据预测市场和订单，优化库存和生产计划、排程。

2.1.3 服务快速反应

通过设备的智能化，可以通过互联网获取用户的实时工况数据。当用户设备出现问题或异常时，帮助用户更快地发现问题、找到问题的原因。通过数据分析，构建基于规则或案例的故障预测系统，对用户设备状态进行预测、帮助用户更好地维护设备。

2.1.4 推动精准营销

利用工业大数据，可以分区域实现对市场波动、宏观经济、气象条件、营销活动、季节周期等多种数据进行融合分析，对产品需求、产品价格等进行定量预测。同时，可以结合用户当前对产品使用的工况数据，对零部件损坏进行预判，进而对零部件库存进行准确调整。此外，通过对智能产品和互联网数据的采集，针对用户使用行为、偏好、负面评价进行精准分析，有助于对客户群体进行分类画像，可以在营销策略、渠道选择等环节提高产品的渗透率。更重要的是，可以结合用户分群实现产品的个性化设计与精准定位，针对不同的群体，实现从产品设计开始实现完整营销环节的精准化。

2.2 促进企业升级转型

随着工业互联网的深入发展，数据集成从企业内部发展到企业之

间。业务应用也随之拓展至终端用户、全产业链和制造服务等场景。这种变化可能导致企业业务定位、盈利模式的重大改变，甚至会导致核心业务的转型。这也就对应了《工业互联网体系架构》中总结的工业智能化四类典型场景：个性化定制、智能化生产、网络化协同、服务化延伸。

2.2.1 创新研发设计模式，实现个性化定制

应用工业互联网和大数据技术，可有效促进产品研发设计的数字化、透明化和智能化。数字化能有效提升效率，透明化可提高管理水平、智能化可降低人的失误。通过对互联网上的用户反馈、评论信息进行收集、分析和挖掘，可挖掘用户深层次的个性化需求。通过建设和完善研发设计知识库，促进数字化图纸、标准零部件库等设计数据在企业内部的知识重用和创新协同，提升企业内部研发资源统筹管理和产业链协同设计能力。通过采集客户个性化需求数据、工业企业生产数据、外部环境数据等信息，建立个性化产品模型，将产品方案、物料清单、工艺方案通过制造执行系统快速传递给生产现场，进行设备调整、原材料准备，实现单件小批量的柔性化生产。

2.2.2 建立先进生产体系，支撑智能化生产

生产过程的智能化是智能制造的重要组成部分。要推进生产过程的智能化，需要对设备、车间到工厂进行全面的数字化改造。以下四点应该特别引起重视：

“数据驱动”。定制化（小批量生产，个性化单件定制）带来的是对生产过程的高度柔性化的要求，而混线生产也成为未来工业生产的一个基本要求。于是，产品信息的数字化、生产过程的数字化成为一个必然的前提。为此，需要为产品相关的零部件与原材料在赛博空间中建立相对应的数字虚体映射，并根据订单与生产工艺信息，通过生产管理系统与供应链和物流系统衔接，驱动相应物料按照生产计划（自动的）流动，满足混料生产情况下物料流动的即时性与准确性要求，从而满足生产需要。

“虚实映射”。个性化或混线生产时，每个产品的加工方式可能是不一样的。这样，当加工过程中的物料按计划到达特定工位时，相应工序的加工工艺和参数（包括工艺要求，作业指导书，甚至三维图纸的信息等等）必须随着物料的到达即时准确的传递到相应的工位，以指导工人进行相应的操作。更进一步的情况下，通过 CPS，生产管理系统将根据这些信息控制着智能化的生产设备自动的进行加工。为此，必须实现数据的端到端集成，将用户需求与加工制造过程及其参数对应起来。同时，通过工业物联网自动采集生产过程和被加工物料的实时状态，反馈到赛博空间，驱使相关数字虚体的对应变化，实现虚实世界的精准映射与变化。

“实时监控”。生产过程及设备状态必须受到严格的监控。当被加工的物料与生产过程中的设备信息在赛博空间实现精准映射的状态下，便可以实现生产过程或产品质量的实时监控。当发现生产过程中出现设备、质量等问题时，便可以及时地通过人或者系统的手段进行

及时的处理。对于无人化、少人化车间，还可以通过网络化的智能系统做到远程监控或移动监控。而要做到这一点，实现正生产过程全流程的纵向集成，便成为必要的前提条件。

“质量追溯”。从订单到生产计划，到产品设计数据，到完整的供应链与生产过程，完整的数据将为生产质量的追溯提供必要的数据保证。信息化系统可以提供订单、供应链与生产计划的完整数据，工业物联网实现了设备、产品与质量数据的采集与存储。这些数据除了保证生产过程的顺利进行，也为未来生产过程的追溯与重现提供了数据基础。为了保证产品质量的持续改进，就要实现从订单到成品的端到端的系统完整信息集成，对生产过程中人机料法环等因素进行准确记录，并与具体订单及相关产品对应，这些是实现完整质量追溯的前提。而系统数据的整合与互联互通，以及不同系统之间数据的映射与匹配，则是实现这个目标的关键所在。

在此基础上，如果能够推进设备的智能化、不断地消除哑设备，通过对积累沉淀的工业大数据的深入挖掘，不断的推进设备与生产过程控制的持续优化，做到设备的自诊断、预测性维护，则对提高设备运行效率、降低维修维护成本、提高产品质量都有着重大意义。

2.2.3 基于全产业链大数据，实现网络化协同

工业互联网引发制造业产业链分工细化，参与企业需根据自身优劣势对业务进行重新取舍。基于工业大数据，驱动制造全生命周期从设计、制造到交付、服务、回收各个环节的智能化升级，推动制造全

产业链智能协同,优化生产要素配置和资源利用,消除低效中间环节,整体提升制造业发展水平和世界竞争力。基于设计资源的社会化共享和参与,企业能够立足自身研发需求开展众创、众包等研发新模式,提升企业利用社会化创新和资金资源能力。基于统一的设计平台与制造资源信息平台,产业链上下游企业可以实现多站点协同、多任务并行,加速新产品协同研发过程。对产品供应链的大数据进行分析,将带来仓储、配送、销售效率的大幅提升和成本的大幅下降。

2.2.4 监控产品运行状态和环境,实现服务化延伸

在工业互联网背景下,以大量行业、用户或业务数据为核心资源,以获取数据为主要竞争手段,以经营数据为核心业务,以各种数据资源的变现为盈利模式,可有力推动企业服务化转型。首先要对产品进行智能化升级,使产品具有感知自身位置、状态能力,并能通过通信配合智能服务,破除哑产品。企业通过监控实时工况数据与环境数据,基于历史数据进行整合分析,可实时提供设备健康状况评估、故障预警和诊断、维修决策等服务。通过金融、地理、环境等“跨界”数据与产业链数据的融合,可创造新的商业价值。例如可通过大量用户数据和交易数据的获取与分析,识别用户需求,提供定制化的交易服务;建立信用体系,提供高效定制化的金融服务;优化物流体系,提供高效和低成本加工配送服务;可通过与金融服务平台结合实现既有技术的产业化转化,实现新的技术创新模式和途径。

2.3 促进中小企业创新创业

互联网促进了协同、降低了交易成本，促进了社会分工。这种逻辑，也同样适用于制造业。

在工业互联网时代，企业有条件服务于更多的客户；故而可以将自己的业务聚焦于更小的细分领域，如特种设备维修维护、特殊产品的制造、专业设计研发等。这样就产生了更多的创业机会。在这种细分市场，中小企业可以通过知识和数据的积累、提升自己的专业水平。

中小企业往往人才和资源匮乏，影响产品创新和生产技术的提升。在工业互联网时代，中小企业可以更加方便地分享和利用外部的人才和知识资源，提升自己的技术水平和创新能力。

工业互联网产业联盟
Alliance of Industrial Internet

3 工业大数据的实施策略

工业大数据的实施策略可以总结为“业务牵引、技术推动、人才支撑”。

3.1 业务的加减乘除

工业大数据作为工业企业实现优化、升级、转型的技术手段，在其实施过程中，企业决策层首先需要考虑的是如何制定业务发展目标。

成功的工业大数据应用实践，大都首先从战略业务目标开始思考，而且往往是由企业领导人亲自驱动，用以终为始的态度进行推进。结合业务总体目标进行逐级分解，梳理出相关的业务方向，进行智能制造中若干环节的优化，亦或是实现某方面智能化水平的升级，再或是拓展企业的业务边界，完成互联网化转型。

在智能化水平提升以及工业互联网转型发展方面，我们把工业企业的相关可量化业务目标总结为加、减、乘、除。业务的加法主要可以从两个维度进行归纳，一个是设计、研发、生产等环节的提质增效，另一个是产品自身的提升，包括产品质量、产品智能化水平等产品竞争力的提升，以及产品业务形态的拓展，例如从卖硬件产品到系统解决方案。减法主要指对现有业务的优化，例如无人少人化、缩短研发周期、降低能耗、压缩服务时间等。乘法对于制造业企业意味着工业互联网平台战略的成功实施，实现用户、零部件供应商、服务商、设计院等外部资源在平台上的聚集、发展与协同，实现企业的指数型发

展。除法是指企业精确定位自己的核心竞争力以及在工业互联网体系中的对应位置和分工，剥离非核心资产，外包高成本业务，实现轻资产化运营。

3.2 过程的智能再造

在明确了业务目标后，确保工业大数据支撑的业务目标的实现不仅仅是一个技术问题，同时企业管理层需要考虑如何实现业务过程再造和相应的组织调整。

首先，企业管理层需要厘清现有业务流程和制造过程，对应明确的业务目标，找到过程中需要改善或者重构的环节。其次，针对这些环节的问题分析，确定相应的解决手段，特别是自动化、信息化与依靠大数据驱动的智能化的分别。再次，业务团队需要定义再造后流程的大致轮廓，并且梳理清楚大数据产生的分析结果与业务流的映射、交互关系以及使用场景。最后，针对新的业务流程，管理层需要制定对应的组织和人员调整预案，确保过程的智能再造能够真正落实。

3.3 架构的统筹规划

工业大数据的实施，需要工业企业信息化部门针对业务问题牵引，结合新型信息技术，特别是大数据的推动，这其中最为重要的是全盘规划设计工业大数据的架构。

工业企业信息化进程参差不齐，工业 2.0、3.0 和 4.0 并存，不同行业之间的数据来源与使用场景千差万别，很难有一个统一的方法来

规划工业大数据架构，但有四方面较为共性问题需要思考。第一，数据资产梳理。意味着企业首先要盘点自有数据资产，包括数据来源、数据规模、数据类型、存储形式等数据基础信息，数据格式、数据质量等问题以及相应数据治理方案，数据管理集成方法，需要从第三方或者互联网获取的外部数据。第二，基础架构选型。需要针对不同业务类型，考虑工作负载特点、数据安全、传输带宽、软硬件成本、运维能力等多方面要素，因地制宜地确定公有云、私有云或混合云等形式的计算平台。第三，原有遗留系统定位。大数据平台不是信息化的替代品，如果说信息化系统重点解决的问题是业务数据的一次传递，那么工业大数据重点解决的问题是工业数据的再次深度利用，大数据应用的数据获取以及分析结果的使用还要与现有信息化系统进行有机融合与结合，所以要能够在规划阶段明确哪些原有系统将被升级替代，哪些将同时存在，以及相应的应用集成、数据集成和分析结果的使用方式。第四，建设开发方法。工业企业根据自身的具体信息化发展水平和人才储备情况，需要在企业大数据平台和大数据应用的建设顺序，以及是自己做还是找外脑等方面的不同方案中做出选择。

3.4 分析算法与模型的领域化

工业大数据分析算法与模型方面应重点关注如下问题：

第一，分析算法与模型的专业化。针对机器时间序列、时空等物联网数据，应提供丰富的特征模板库，方便对典型物理事件（如风速平稳时段、发电机转速快速下降、环境温度逐渐上升等）进行描述；

另外，还应提供丰富的时间序列、时空模式、序列模式的深度挖掘算法库，提升工业数据分析能力与效率。

第二，跨时间尺度数据综合分析。与传统商务智能（**Business Intelligence, BI**）应用场景不同，工业大数据包括秒/小时尺度的机器工况数据、天/周尺度的车间调度数据和月/年尺度的管理决策数据，在工业大数据应用中不同时间尺度的工业数据需要综合使用才能驱动智慧企业。比如，智慧矿山企业需要根据跨界数据预测矿石品质需求，依据品质需求生成装备作业调度计划，根据装备作业计划实时控制装备操作。

第三，交互式分析。工业大数据分析必须能够满足大规模、分散控制和交互迭代等需求。在实时处理上，传统的商业数据分析系统不能有效支持面向大规模数据状态下的低时延复杂事件检测。在离线分析上，查询检索与分析建模应紧密协同，同时前台探索展现与大数据深度分析要无缝整合，支持领域专家交互挖掘。

第四，领域知识自动化。工业大数据中存在着大量的非结构化数据，将当前深度学习、自然语言处理等人工能力融入到工业大数据分析中来，可以有效辅助把专家知识进行有效的沉淀、萃取和自动化。

3.5 人才的培养引进

工业大数据的实施和应用不是一个交钥匙的工程，基于大数据的创新更离不开专业人才。除机械和电子工程等技术能力以外，还需要一批新兴的跨学科人才。比如下一代工程学人才，将机械工程和信息

化融汇贯通，成为“数据机械工程师”。此外，还需要培养数据科学家、用户界面专家等跨界人才。目前中国制造业企业因为行业领域区域等制约，大数据分析人才难以满足需求。为此，企业自身需要制定复合型人才培养计划，营造氛围提升员工的综合数据处理能力，另一方面需要创造环境积极引进大数据分析人才，同时探索“不求所有、但为所用”的新型人才队伍建设模式。



工业互联网产业联盟
Alliance of Industrial Internet

4 工业大数据系统与技术架构

4.1 工业大数据生命周期

基于工业互联网的网络、数据与安全，工业大数据将构建面向工业智能化发展的三大优化闭环处理流程。一是面向机器设备运行优化的闭环，核心是基于对机器操作数据、生产环境数据的实时感知和边缘计算，实现机器设备的动态优化调整，构建智能机器和柔性产线；二是面向生产运营优化的闭环，核心是基于信息系统数据、制造执行系统数据、控制系统数据的集成处理和大数据建模分析，实现生产运营管理的动态优化调整，形成各种场景下的智能生产模式；三是面向企业协同、用户交互与产品服务优化的闭环，核心是基于供应链数据、用户需求数据、产品服务数据的综合集成与分析，实现企业资源组织和商业活动的创新，形成网络化协同、个性化定制、服务化延伸等新模式。

工业大数据的处理过程符合大数据分析生命周期，涉及多个不同阶段，如下图所示：

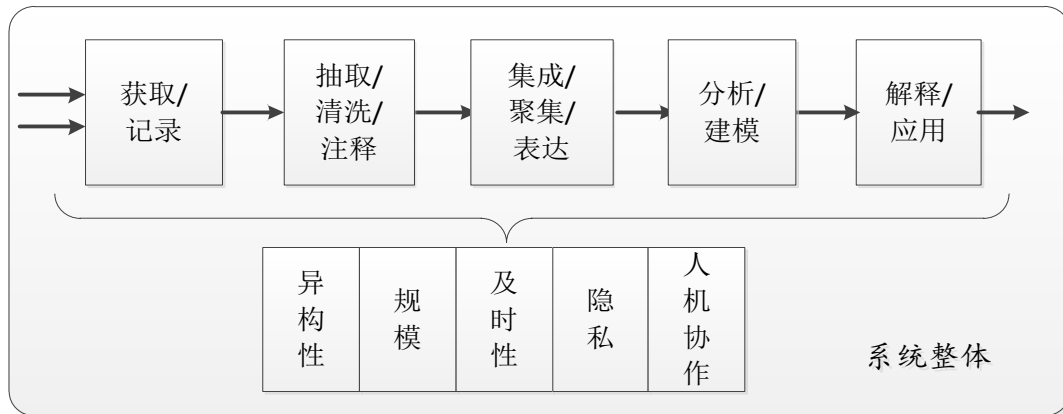


图1 工业大数据生命周期

大数据生命周期的主要环节如图的上半部分所示，图的下半部分是大数据分析的需求，这些需求使得分析任务具有挑战性。

工业大数据数据采集阶段重点关注如何自动地生成正确的元数据以及其可追溯性。既要研究如何生成正确的元数据，又要支持进行数据溯源。

工业大数据抽取、清洗和注释阶段主要负责对工业数据集进行数据抽取、格式转换、数据清洗、语义标注等预处理工作，是数据工程的主要内容。

工业大数据集成、聚集与表示阶段主要关注数据源的“完整性”，克服“信息孤岛”，通常工业数据源通常是离散的和非同步的。对于飞机、船舶等具有复杂结构的工业产品，基于**BOM**进行全生命周期数据集成是被工业信息化实践所证明的行之有效的方法。对于化工、原材料等流程工业产品，则一般基于业务过程进行数据集成。

工业大数据建模和分析阶段必须结合专业知识，工业大数据应用强调分析结果的可靠性，以及分析结果能够用专业知识进行解释。工业大数据是超复杂结构数据，一个结果的产生，是多个因素

共同作用的结果，必须借助专业知识，同时，工业过程非常复杂，现实中还可能存在很多矛盾的解释，因此，要利用大数据具有“混杂”性的特点，通过多种相对独立的角度来验证分析结果。

工业大数据分析结果解释与应用阶段要面对具体行业和具体领域，以最易懂的方式，向用户展示查询结果。这样做有助于分析结果的解释，易于和产品用户的协作，更重要的是推动工业大数据分析结果闭环应用到工业中的增值环节，以创造价值。

4.2 工业大数据技术架构

参考工业互联网架构，本节从功能层面给出工业大数据的技术架构。其中工业大数据管理能力包含数据采集与交换、数据预处理与存储、数据工程与数据建模四部分，工业大数据分析能力涵盖支持离线批量计算和在线实时计算的分布式分析框架，以及工业领域分析会使用到的各类分析算法库。

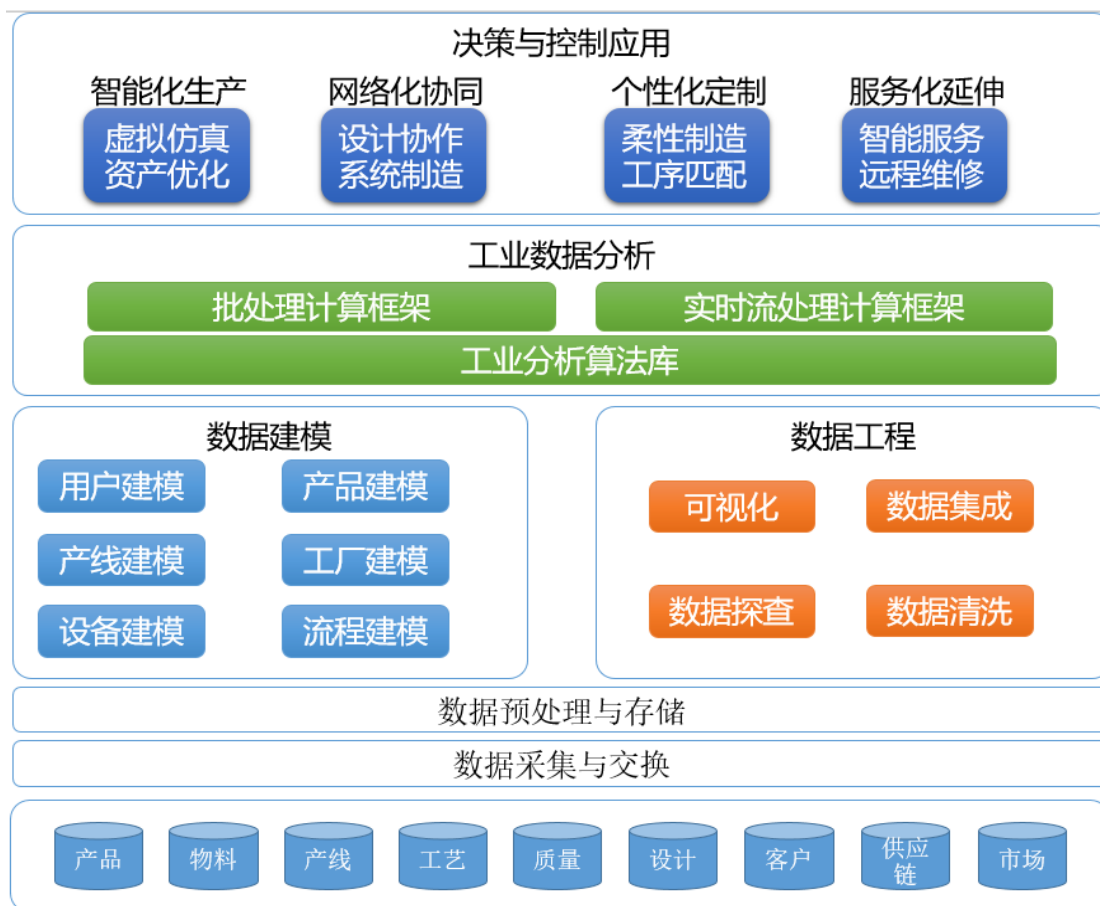


图 2 工业大数据技术架构

4.3 工业大数据管理技术架构

从功能上来说，工业大数据管理技术分为四部分：数据采集与交换、数据预处理与存储、数据工程与数据建模。

数据采集与交换层主要负责通过 PLC、SCADA、DCS 等系统从机器设备实时采集数据，也可以通过数据交换接口从实时数据库等系统以透传或批量同步的方式获取物联网数据。同时该层提供数据交换接口，可以从业务系统的关系型数据库、文件系统中采集所需的结构化与非结构化业务数据，也可反向将汇总、分析结果等数据传回业务系统。此外，该层还提供与外部系统的数据交换、互联网数据爬取、

第三方数据服务访问的能力。

数据预处理与存储层主要负责对采集到的数据进行数据解析、格式转换、元数据提取、初步清洗等预处理工作，再按照不同的数据类型与数据使用特点选择分布式文件系统、NoSQL 数据库、关系数据库、对象存储系统、时序数据库等不同的数据管理引擎实现数据的分区选择、落地存储、编目与索引等操作。

数据工程层主要完成对工业大数据的治理并支撑对数据的探索能力，以供应用开发与分析对数据的方便使用。其包括方便用户了解数据特性并发现数据潜在问题的数据探查能力、对有质量问题数据的清洗与修正能力、对异构数据的集成能力以及对各类工业数据的可视化展现能力。

数据模型层主要完成对底层数据模型的工业语义封装，构建基于用户、产线、工厂、设备、产品等对象的统一数据模型，对各类统计分析应用与用户实现更加便捷、易用的数据访问接口。

4.4 工业大数据分析技术架构

从功能上来说，工业大数据分析技术分为两大部分：分布式计算框架与工业大数据分析算法库。

分布式计算框架主要负责对分析在线实时分析任务与离线批量数据分析任务的调度与执行，特别是针对大数据的分布式数据密集型计算。同时对着深度学习技术的发展，对于深度学习框架的支持，GPU 等新型硬件的计算加速也是必不可少的能力。

工业大数据分析算法库除了典型的机器学习算法模型外，需要针对工业特有的稳态时间序列、时空等数据提供丰富的特征模板库，方便对典型物理事件在时域和频域上的精确描述；另外，还应提供丰富的时间序列、时空模式、序列模式的深度挖掘算法库。



工业互联网产业联盟
Alliance of Industrial Internet

5 工业大数据管理技术

本章主要针对工业大数据多样性、多模态、高通量和强关联等特性，重点阐述工业大数据管理中主要技术。

5.1 多样性数据的采集技术

工业软硬件系统本身具有较强的封闭性和复杂性，不同系统的数据格式、接口协议都不相同，甚至同一设备同一型号的不同时间出厂的产品所包含的字段数量与名称也会有所差异，因而无论对于采集系统对数据进行解析，还是后台数据存储系统对数据进行结构化分解都会带来巨大的挑战。由于协议的封闭，甚至无法完成从设备对数据的采集，在可以采集的情况下，在一个工业大数据项目实施过程中，通常也至少需要数月的时间对数据格式与字段进行梳理。挑战性更大的是多样性的非结构化数据，由于工业软件的封闭性，数据通常只有特定软件才能打开，从中提取更多有意义的结构化信息工作通常很难完成。这类挑战需要工业标准化的推进与数据模型自动识别、匹配等大数据管理技术进步共同解决。

5.2 多模态数据的管理技术

各种工业场景中存在大量多源异构数据，例如结构化业务数据、时序的设备监测数据、非结构化工程数据等。每一类型数据都需要高效的存储管理方法与异构的存储引擎，但现有大数据技术难以满足全

部要求。以非结构化工程数据为例，特别是对海量设计文件、仿真文件、图片、文档等小文件，需要按产品生命周期、项目、BOM 结构等多种维度进行灵活有效的组织、查询，同时需要对数据进行批量分析、建模，对于分布式文件系统和对象存储系统均存在技术盲点。另外从使用角度上，异构数据需要从数据模型和查询接口方面实现一体化的管理。例如在物联网数据分析中，需要大量关联传感器部署信息等静态数据，而此类操作通常需要将时间序列数据与结构化数据进行跨库连接，因而需要针对多模态工业大数据的一体化查询协同进行优化。

5.3 高通量数据的写入技术

在越来越多工业信息化系统以外的数据被引入大数据系统的情况下，特别是针对传感器产生的海量时间序列数据，一个装备制造企业同时接入的设备数量可达数十万台，数据的写入吞吐达到了百万数据点/秒-千万数据点/秒，大数据平台需要具备与实时数据库一样的数据写入能力。考虑到大数据平台要对数据进行长时间存储，其高效的数据编码压缩方法以及低成本的分分布式扩展能力也是重要的挑战。另一方面，数据在使用上，不仅是对数据在时间维度进行简单地回放，而且对于数据多条件复杂查询以及分析性查询也有着极高的要求。因此，针对数据写入面临的挑战，工业大数据平台需要同时考虑面向查询优化的数据组织和索引结构，并在数据写入过程中进行一定的辅助数据结构预计算，实现读写协同优化的高通量数据写入。

5.4 强关联数据的集成技术

工业大数据分析更关注数据源的“完整性”，而不仅仅是数据的规模。由于“信息孤岛”的存在，这些数据源通常是离散的和非同步的。工业大数据应用需要实现数据在物理信息、产业链、以及跨界三个层次的融合。

“物理信息融合”：表现在设计开发阶段主要管理数字产品，而在制造服务阶段主要管理物理产品，跨生命周期管理需要融合数字产品和物理产品，从而构建工业信息物理融合系统。

“产业链融合”：表现在互联网大数据环境下，以资源整合优化为目标的云制造模式得以迅速发展，智能产业链需要突破传统企业边界，实现数据驱动的业务过程集成。

“跨界融合”：是指在“互联网+”环境下，企业需要将外部跨界数据来源进行集成，比如，美国某农机公司将天气数据、灌溉数据、种子数据以及农机数据进行综合利用，为农场提供粮食增产服务。

这与其它领域大数据集成具有明显差异，因此不仅需要从数据模型，更需要从制造过程、层次化物料表(Bill Of Material, BOM)结构、运行环境等多类型工业语义层面对工业大数据进行一体化整合管理，其中 **BOM** 是产品全生命周期数据集成的关键手段。**BOM** 定义了企业信息系统数据的核心语义结构。针对装备物联网数据和外部互联网数据,可以根据其绑定的物理对象(零部件或产品)与相应的 **BOM** 节点相关联。从而以 **BOM** 为桥梁来关联三个不同来源的工业大数据。

具体实现机制可以分为三个层面，逻辑层负责统一数据建模，定义数字与物理对象模型，完成底层数据模型到对象模型映射；概念层实现数据语义层面的融合，通过语义提取与语义关联，形成 **RDF** 形态的知识图谱，提供基于 **SPARQL** 的查询接口；操作执行层负责异构数据管理引擎的查询协同优化，对外提供 **SQL** 以及 **REST API** 形式的统一查询接口。



工业互联网产业联盟
Alliance of Industrial Internet

6 工业大数据分析技术

本章针对工业数据的强机理、低质量和高效率要求，重点阐述工业大数据分析主要技术。

6.1 强机理业务的分析技术

工业过程通常是基于“强机理”的可控过程，存在大量理论模型，刻画了现实世界中的物理、化学、生化等动态过程。另外，也存在着很多的闭环控制/调节逻辑，让过程朝着设计的目标逼近。在传统的数据分析技术上，很少考虑机理模型（完全是数据驱动）、也很少考虑闭环控制逻辑的存在。强机理对分析技术的挑战主要体现在三个方面：

- 1) 机理模型的融合机制，如何将机理模型引入到数据模型(比如机理模型为分析模型提供关键特征，分析模型做机理模型的后处理或多模型集合预测)或者将数据模型输入到机理模型（提供 **parameter calibration**）；
- 2) 计算模式上融合，机理模型通常是计算密集型（CPU 多核或计算 **cluster** 并行化）或内存密集型（GPU 并行化），而数据分析通常是 I/O 密集型（采用 **Map-reduce**、**parameter server** 等机制），二者的计算瓶颈不同，需要分析算法甚至分析软件需要特别的考虑；
- 3) 与领域专家经验知识的融合方法，突破现有生产技术人员知识盲点，实现过程痕迹的可视化。比如，对于物理过程环节，重视知识的“自动化”，而不是知识的“发现”。将领域知识进行系统化管理，通过大数据分析进行检索和更新优化；对于相对明确的专家知识，借助

大数据建模工具提供的典型时空模式描述与识别技术，进行形式化建模，在海量历史数据上进行验证和优化，不断萃取专家知识。

6.2 低质量数据的处理技术

“大数据分析”期待利用数据规模弥补数据的低质量。由于工业数据中变量代表着明确的物理含义，低质量数据会改变不同变量之间的函数关系，给工业大数据分析带来灾难性的影响。但事实上制造业企业的信息系统数据质量仍然存在大量的问题，例如 ERP 系统中物料存的“一物多码”问题。物联网数据质量也堪忧，无效工况（如盾构机传回了工程车工况）、重名工况（同一状态工况使用不同名字）、时标错误（如当前时间为 1999 年），时标不齐（PLC 与 SCADA 时标对不上）等数据质量问题在很多真实案例中可以达到 30% 以上。这些质量问题都大大限制了对数据的深入分析，因而需要在数据分析工作之前进行系统的数据治理。

工业应用中因为技术可行性、实施成本等原因，很多关键的量没有被测量、或没有被充分测量（时间/空间采样不够、存在缺失等）、或者没有被精确测量（数值精度低），这就要求分析算法能够在“不完备”、“不完美”、“不精准”的数据条件下工作。在技术路线上，可大力发展基于工业大数据分析的“软”测量技术，即通过大数据分析，建立指标间的关联关系模型，通过易测的过程量去推断难测的过程量，提升生产过程的整体可观可控。

6.3 数据高效率处理技术

工业大数据分析在分析过程和运算的效率方面主要面临两个技术难点。

一是工业应用特定数据结构带来的新需求。通用的数据分析平台大多针对记录性数据或独立的非结构化数据（适合交易、业务运营管理、社交媒体等场景）。然而工业应用常常依赖于大量时序或时空数据（传感器数据）和复杂的产品结构（如层次性的离散 BOM 结构、或线性连接结构），这就需要工业大数据分析软件在底层数据结构设计、基础分析算法和建模过程上，提供充分支持，例如，复杂 BOM 结构的离散装备的分析建模、多变量非线性时间序列特征提取与处理算法（信号分解、降噪、滤波、序列片段切割）等。

二是工业应用模式对分析处理效率的要求。工业应用模式通常是大规模分布式（空间）、实时动态（时间）、异构性强（连接）。这对分析平台软件提出了新的挑战。在实时处理上，需要能够支持面向大规模数据状态下的低等待时间复杂事件检测。在离线分析上，前台分析建模应与后台的工业大数据平台应有很好的整合，支持大数据上的挖掘。

7 工业大数据应用案例

7.1 北京工业大数据创新中心推进风电装备数字化升级

基本情况

为贯彻《中国制造 2025》和《中国制造 2025 北京行动纲要》，支持《国家制造业创新建设工程》，在工信部和北京市经信委指导下，北京工业大数据创新中心由清华大学和昆仑数据牵头，联合业界 19 家龙头企业、科研院所及高校组建成立。致力于中国工业大数据平台的自主核心技术突破，昆仑数据推出了国内首款专为机器数据优化的 **KMX** 工业大数据管理分析平台，让工业专家无需掌握大数据技术就可以轻松应用大数据解决业务问题。

北京工业大数据创新中心成员单位新疆金风科技股份有限公司（以下简称“金风科技”）成立于 1998 年，是中国最早从事风电机组研发和制造企业之一。金风科技拥有自主知识产权的直驱永磁技术，代表着全球风力发电领域最具成长前景的技术路线。并已发展成为国内第一、国际市场排名第一的风电机组制造商及风电整体解决方案供应商，产品服务于全球六大洲、17 个国家，占中国出口海外风电机组容量的 50% 以上。

昆仑数据与金风科技于 2014 年开始协作进行风力装备的工业大数据应用探索，从故障预警、运营优化等方面着手挖掘大数据价值。目前，已将风机的设计/仿真数据、运维档案、风机状态监测数据、测

风塔观测数据、气象数据和地理信息等风电数据资源池统一整合到基于 KMX 的风电大数据平台，将通过装备智能化、供应链协同、跨生态整合三条路径，逐步实践风力装备的数字化升级。

业务痛点

我国风电装机虽然增长迅猛，但在风电利用上才刚刚起步。2015 年底，风电占全部发电比重刚刚超过 3%。

1) 存量风场的弃风限电挑战

三北地区风资源丰富但用电企业少，导致能源无法就地消纳；薄弱的网架结构以及特高压交流建设相对缓慢严重制约了跨区域送电能力，弃风限电现象比较严重。除了近一步提升本地能源消纳能力，另一个突破点是运用大数据技术，提升风力发电机组的运行可靠性，在风资源条件具备和并网条件成熟的情况下，避免因设备故障等因素带来设备和发电量损失。

2) 新增市场的开发风险控制

经济发达的东南沿海和南方地区，对风电的消纳能力较好。然而相比三北地区，不管海上风场还是山地风场，其风资源条件更为复杂或相对较差，同时还需综合考虑海洋洋流、海拔地形等多种环境因素，如何保证和提升风机出力效率并降低维护成本面临较大挑战。风场建设的技術要求和经济性风险较高，建设进度缓慢，需要通过精益化控制和无人值守提升风场运营的效率，降低风场建设的经济性风险。

3) 行业竞争面临腹背夹击

业务挑战的另外一方面来自于传统能源的主导地位优势和新能源补贴的逐步降低。十三五期间，风电并网电价和补贴需要在五年内在现有电价基础上下降 20%到 25%，这是国家能源局对风电行业生存发展的硬指标。同时在海外市场上也面临着西门子、通用电气 (GE)和维斯塔斯等跨国企业的竞争压力。

借助大数据和云计算等手段降低建设及运维成本、提高风功率预测精准度、以争取优先调度主动权已经成为风力装备企业的共识。

数据来源

风电作为一种不稳定能源，其开发和应用对数据有很强的依赖性，自动化程度相较于其它工业领域也相对较高。从风场风机的规划设计、风场的建设交付到风场的运维管理，风电数据来源相当广泛。在实际业务中，常常需要将多种数据融合使用。

1) 机器数据

风电行业中的机器数据包括风机运行监测数据、风机运行监控数据、设备故障数据等。主要以时序数据，以数据量大、接入频度高为主要特点。为了更好地监测风机实际运行情况，每台风机的数据测点从最初的十几个增长到如今近五百个，数据回传频率从秒级提高到每秒 50 组的高频，峰值状态下 20000 台风机每秒会产生逾千万条、每天新增容量近 1TB 的传感器数据和机组运行日志数据，并随着装机容量和控制精益化要求的提高，未来的数据量会进一步提升。

2) 外部产业链数据

风能开发强烈依赖于环境，因此环境数据也是重要的数据来源。环境数据可以分为风资源数据、地理信息数据和气象数据。数据类型比较丰富，包括结构化、半结构化和非结构化数据。风资源数据来源于测风塔和激光雷达，数据采集频率从分钟级到小时不等，每年新增数据超过 100G。地理信息数据包括测绘地图、卫星遥感图像、地质属性等，以存量数据为主，数据量近 50TB。气象数据包括气象卫星数据、地面气象站数据、中尺度气象数据等，数据量超过 1PB。

3) 业务数据

通常仿真软件一轮仿真会产生 30GB 的数据，为了在低风速风场、海上风场的复杂工况下确保风场开发的经济性，随着混排技术和精益控制技术的引入，一轮仿真产生的数据会指数级膨胀至 700GB。为了支持上层的数据应用，系统还需要集成融合如服务管理数据、ERP 数据、供应链数据等来自于信息系统的数据。这部分的数据量主要以关系型数据为主，数据总量相对较低，总量约 10TB。

技术方案

KMX 工业大数据平台整体基于自主可控的大数据技术，以分布式微服务体系建造，为今后的扩展性和发展打下了基础。

1) 确保海量高速数据的高效接入和存储

首先 **KMX** 从技术和经济性上确保海量高速机器数据的高效接入和存储问题。**KMX** 通过自适应高通量数据接入服务，持续接收和管理以千万测点每秒的速率产生的机器数据，确保数据的高可用性。

并在数据的入口即进行数据的格式和质量检查，及时发现数据问题，避免垃圾进垃圾出。通过内置多种数据存储引擎，满足结构化、时序与非结构化等异构工业数据的存储要求。基于通用普通商业服务器的线性扩容机制以及针对时序机器数据特有的压缩算法，保证了数据存储的经济性。

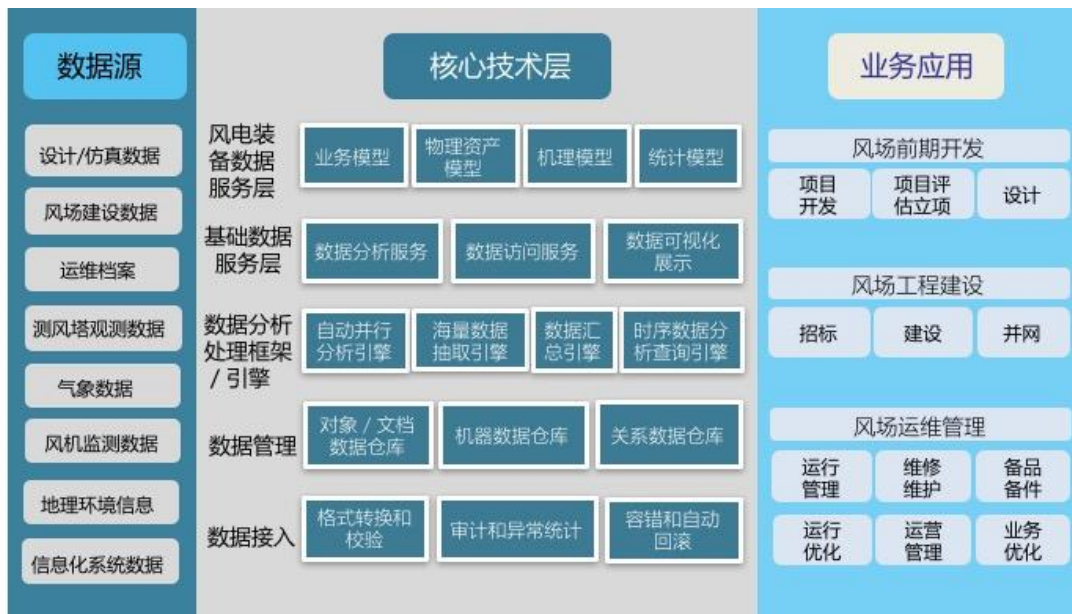


图3 KMX工业大数据平台风电领域应用技术架构

2) 内置工业数据分析查询引擎和行业模型，提升数据价值的高效挖掘能力

系统内置工业数据查询引擎，支持丰富的工业查询语义，百倍提升了数据查询效率，实现了分析查询一体化。针对风电行业业务场景和应用，系统还分别从业务域和物理域内置了业务模型，如运维成本模型、发电收益模型；物理资产模型，如风机的BOM结构、全生命周期模型；以及结构力学模型、风机动力学模型、流场模型等机理模型和统计模型，将系统内的工业数据有机整合，打造数字孪生体，实

时反应并预测物理世界风机的运行状态。结合特有的非侵入式的大数据并行化分析引擎，通过大数据与小数据的无缝衔接，降低大数据分析使用门槛。

3) 打造行业数据分析服务，加速数据价值释放和落地

围绕风资源评估、电网友好性设计、发电量提升、资产寿命延长、资产可靠性提升等风电行业核心业务场景，系统融合业务数据、外部数据和机器数据，运用机器学习、深度学习等数据技术，打造风电行业数据分析服务，提供风电场设计仿真服务、发电资产健康评估服务、故障预警预测服务，快速推进相关业务如风场定制化设计、设备运维管理发展和数字化转型，使业务人员也可以使用大数据轻松应对和解决业务问题，确保数据行业价值的快速落地。

应用效益

经测算，基于 **KMX** 工业大数据平台的风电装备智能化升级方案迄今已产生经济效益达数亿元。

首先实现了发电效率的提升。通过数据分析优化现有风机运行状态,提升风力发电机组运行效率。以对风偏航角的优化为例，对偏航角优化算法并行化处理后，由于对风准确性提升和偏航响应时间的缩短，一台风机一年能多产出 1 万元电量，按市场保有量 2 万台风机换算，一年能提升 2 亿元能效产值。

其次更好地控制了运维成本。通过对全球风电场的实时监控，建立故障预警模型近百项，为“无人值守”模式奠定基础。在 **KMX** 平台

的支撑下，某关键零部件故障预警可以提前 72 小时，通过主动性维修，可以降低 90% 因该部件故障而产生的次生事故，每年减少因此导致的风场直接和间接损失可达千万元。

同时还将大幅提升风电企业的研发和创新效率。例如通过 KMX 的大数据优化处理，将单轮 30G 仿真数据的后处理速度从几个小时降低到了几分钟，极大提升了风场制化设计的迭代速度和开发效率。

随着风电数字化转型整体方案逐步落地，将使风电的度电成本降低 25%，大幅提升不稳定资源利用的可靠性和经济性，从而提升风电的市场竞争力。根据十三五新能源规划，2020 年风电电量占比应从 2015 年的 3% 提升至 6-7%，工业大数据将在我国能源转型中扮演更加重要的角色。

7.2 北科亿力推动炼铁行业大数据应用

基本情况

北京北科亿力科技有限公司由来自北京科技大学、东北大学、国内知名钢铁企业的冶金、自动化、软件等专业人才创建。其创始团队自 2003 年起一直致力于炼铁数字化及智能化技术的研发。公司成立于 2008 年，其最终用户为以炼铁厂为主体，包含设计院所、高校等机构的冶金生态圈。到 2016 年，公司业务已覆盖全类型高炉，完成了全国约 20% 炼铁产能的数字化及智能化项目，并投资 1185 万元用于企业及行业级炼铁大数据平台的建设。

钢铁是国家战略性支柱行业，连续多年占全国 GDP 总值的 10% 以

上，在去产能背景下其体量依然巨大。作为占据钢铁企业约 70% 的成本和能耗的炼铁厂，2016 年中国铁水产量约 7 亿吨，产值约 1.5 万亿，其生产过程的数字化和智能化水平相比于钢轧工序依然较低。通过炼铁大数据平台和智能化系统的建设，降低炼铁异常工况及燃料消耗，不但能为炼铁厂带来巨大的经济效益，还可以直接降低炼铁能耗及 CO₂ 排放，实现节能减排和绿色冶金。

北科亿力已获得炼铁大数据及数字化监测方面的软件著作权 21 项、专利 11 项，自主研发炼铁大数据互联平台推动炼铁智能制造：将无线传输、三维激光雷达、热成像技术应用于高炉“自感知”工业传感及物联网开发；建立了数据源、整合、传输、管理、持久、分析、接口、应用的大数据处理中心；开发高炉专家系统实现自诊断和自决策；建立行业级炼铁大数据平台实现企业端和行业端数据互联互通及智能对标；开发云平台实现移动终端服务。广泛应用于河钢、首钢、中信特钢、沙钢、山钢、酒钢等中国近百家和海外伊朗、越南、印尼等 200 多座高炉，在实际应用中对炼铁长寿、高效、优质、低耗、清洁生产起到了至关重要的作用。

业务痛点

钢铁工业是国家战略性支柱行业，连续多年占全国 GDP 总值的 10% 以上。炼铁工序占据钢铁全流程约 70% 的冶炼成本和能耗。炼铁工艺是透明的，数据是开放的，但智能化发展滞后，其业务痛点在于：

- 1) 炼铁大数据缺乏利用、智能炼铁相对空白

目前炼铁厂已有的信息化系统偏重于基础自动化和生产计划管理，和冶炼过程的智能控制基本脱节，这是由于炼铁核心单元高炉具备高温、高压、密闭、连续的“超大型黑箱”特性，高炉内部工作状态诊断困难，操作仍以人工经验和主观判断为主，中国制造 2025 在智能炼铁领域相对空白。

2) 数字化、标准化炼铁尚未普及，各企业间炼铁成本及能耗水平参差不齐

由于数字化、智能化、标准化炼铁体系尚未建立，行业内各炼铁厂技术经济指标参差不齐，即使是重点钢企，其吨铁能耗相差大的都高达 130 公斤/吨，寿命相差最大的达 18 年，成本相差大的超过 500 元/吨。

3) 炼铁行业数据共享壁垒造成数据交互和技术推广困难，行业可提升空间巨大

行业级炼铁大数据共享缺少平台，数据交互、咨询诊断、技术推广及案例共享困难，缺少对行业级海量数据的深度分析、挖掘和利用，行业级大数据平台的体系化建设有巨大的发展空间：从整个炼铁行业而言，2016 年中国铁水产量 7 亿吨，产值近 1.5 万亿，基于行业级大数据的互联互通实现整体提升，降低吨铁成本和能耗，每年至少有 70 亿元的创效空间及 1000 万吨的 CO₂ 减排潜力。

4) 炼铁行业生态圈发展滞后造成资源共享效率低下

对于整个炼铁生态圈而言，设计、生产、科研、标准、管理、供应等相互之间仍存在信息壁垒，无法整合炼铁上、中、下游的纵向资源，

以及与炼铁相关的横向资源，给整体炼铁生态和行业的发展造成很大障碍。

数据来源

炼铁大数据平台主要数据来源包括：

1) 物联网机器数据

主要包括炼铁 PLC 生产操作数据、工业传感器产生的检测数据、现场各类就地仪表的数据等。整个炼铁大数据平台目前已接入了约 200 座高炉的数据，以单座高炉为例，每个高炉约有 2000 个数据点，数据采集频率为 1 分钟一次，每座高炉产生的数据量约为 288 万点/天、数据大小约为 200Mb/天，即行业大数据平台接入的数据量约为 5.76 亿点/天，数据大小约为 40G/天。

2) 内部核心业务数据

主要包括 LINES 系统的检化验数据、MES 系统的生产计划数据、DCS 系统的过程控制数据、ERP 系统的成本设备数据、用户的交互需求数据、模型计算及分析结果形成知识库的数据、以及现场实际生产过程中的经验数据信息等。200 座高炉的相关数据整合到炼铁大数据平台后形成 TB 级的数据信息。

3) 外部应用平台数据

主要包括国家和行业标准、电子期刊、专家知识库、数据案例和相关政策信息等。通过购买、互联网收集、用户提供等，形成 TB 级的数据量，并且进行实时更新。

技术方案

炼铁大数据平台通过在企业端部署自主研发的工业传感器物联网，对高炉“黑箱”可视化，实现了企业端“自感知”；通过数据采集平台将实时数据上传到大数据中心；通过分布式计算引擎等对数据进行综合加工、处理和挖掘；在业务层以机理模型集合为核心，结合多维度大数据信息形成大数据平台的核心业务，包括物料利用模块、安全预警模块、经济指标模块、工艺机理模块、精细管理模块、智能生产模块、设备监管模块、资产管理模块、数据样本模块、能耗监控模块等；应用传输原理、热力学动力学、炼铁学、大数据、机器学习等技术建立高炉专家系统，结合大数据及知识库，实现“自诊断”、“自决策”和“自适应”。

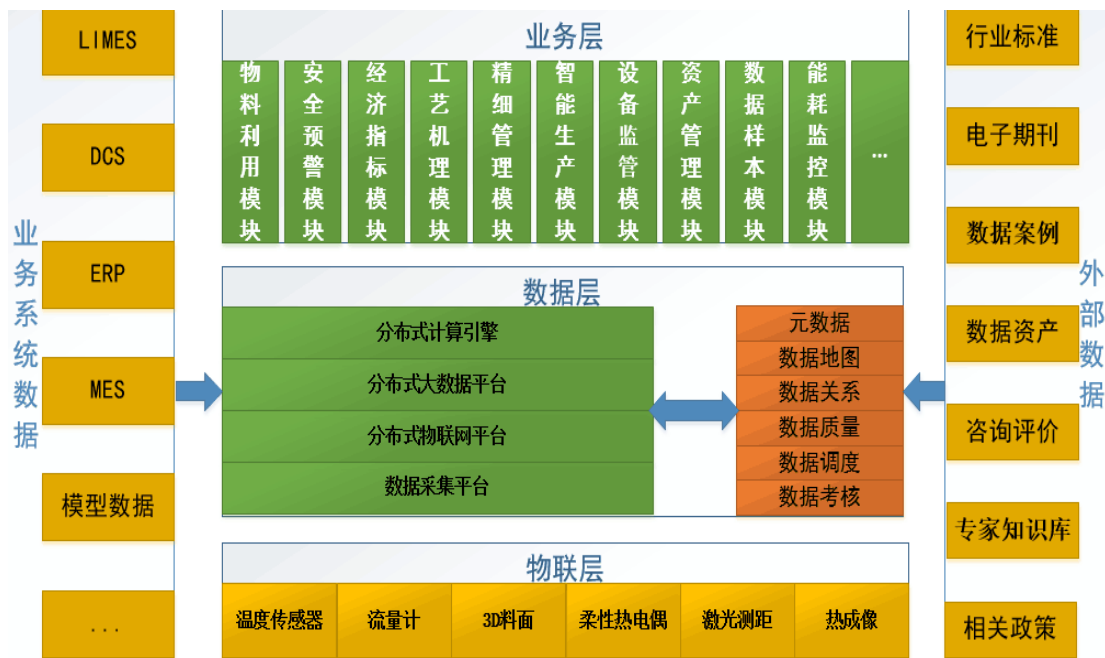


图 4 北科亿力炼铁大数据技术架构

通过推行炼铁物联网建设标准化、炼铁大数据结构和数据仓库标

准化、数字化冶炼技术体系标准化，建立行业级炼铁大数据智能互联平台，实现各高炉间的数据对标和生产优化，促进设计院、学会、供应商、科研机构等整个生态圈的信息互联互通、数据深度应用、产学研用紧密结合和核心竞争力的提高。

1) 自主知识产权的炼铁智能监测硬件及物联网已覆盖全国约 20% 炼铁产能

自主研发的炼铁工业传感器、监测系统及物联网已覆盖全国约 20% 炼铁产能，已投入 4.5 亿元在各地高炉部署温度传感器约 35000 支、流量计 4000 台、热电偶 30000 支、数学模型系统 200 套以上，实时自动采集数据，确保数据真实性。通过在现场布置各种炼铁传感器并联网，为机理模型和大数据挖掘及智能诊断提供自动、实时、真实的基础数据。

2) 深厚的机理研究、全系列的数学模型及成功的应用实践

研发团队以冶炼专业为主结合大数据、人工智能、软件工程等专业，通过 20 年以上机理研究，发表专业文章 100 余篇，获得知识产权 32 项。根据传输原理、物理化学、炼铁学建立炼铁物料、能量、安全、冶炼机理等全系列的 35 个数学模型，并提炼无量纲的准数，从安全、节能、高效生产、KPI 管理等不同层面建立数据对比方法和有效的评价标准。机理模型及数字化炼铁技术为单座高炉实际创效 2400 万元/年，并获省部级科技进步奖多项。

3) 大数据与机器学习相结合，建立知识库和推理机，推行数字化、智能化、标准化炼铁

基于冶炼机理的无量纲指标实现不同企业、不同类型高炉间的数据对标，通过大数据挖掘和云诊断、机器学习相结合，总结 3000 多条规则并建立炼铁工况知识库和推理机。将各个炼铁现场海量的监控系统、检化验系统、生产运营系统的基础数据汇聚并建立分级数据仓库进行存储并支持快速检索，对海量数据进行深度挖掘，结合机理模型及核心评价标准，对各企业、各工序、各人员操作数据进行横向及纵向对比分析，推行数字化、智能化、标准化炼铁。

应用效益

1) 经济效益

通过炼铁大数据智能互联平台的建立，提升炼铁的数字化、智能化、科学化、标准化水平，预判和预防高炉异常炉况的发生，提高冶炼过程热能和化学能利用效率，已应用的炼铁厂平均提高劳动生产率 5%，降低冶炼燃料比 10 公斤/吨铁，降低吨铁成本 15 元，直接经济效益单座高炉创效 2400 万元/年；预期全行业推广后，按中国 7 亿吨/年的铁水产能，吨铁成本降低 10 元计，直接经济效益 70 亿元/年。

2) 社会效益

已应用的炼铁厂减少 CO₂ 排放 10 公斤/吨铁；预期全行业推广后 CO₂ 减排 1050 万吨/年。炼铁大数据应用助力绿色冶金，实现低燃料消耗和节能减排，减轻炼铁和炼焦造成的环境污染。

7.3 大唐集团工业大数据应用

基本情况

中国大唐集团公司（以下简称“大唐集团”）是在电力体制改革中组建的中央直接管理的特大型发电企业。主要从事电力、热力生产和供应，与电力相关的煤炭资源开发和生产，以及相关专业技术服务，重点涉及发电、供热、煤炭、煤化工、金融、物流、科技环保等业务领域。2010年首次入围世界500强居412位，2016年居第406位。装机容量和发电量均占全国的9%左右。

大唐集团自2012年起，耗费近3年时间研究了大数据、云计算、物联网、人工智能等关键技术，开发了大数据建模、实时数据库、数据甄别等领域新产品，建立了集团数据共享和可视化分析中心，实现了科学预测与智能决策等全集团全过程全要素的数据深度分析与挖掘，充分发挥数据的价值，以数据说话，以数据决策。

业务痛点

发电集团企业资产遍布全国，经营区域广、资产规模大、管理链条长、技术含量高，传统的管理手段已经不能满足集团整体管控能力的要求，尤其体现在生产管理、燃料管理等方面。

1) 生产管理不集中，实时监控能力较弱

发电集团主要是为电网提供清洁电力，对电力生产过程安全可靠、经济环保有很高的要求。生产实时监控能力较弱，非停事故预防能力不强，早期状态诊断缺乏有效手段，优化运行分析指导缺乏有效平台

是大部分发电企业面临的主要问题。

2) 燃料管理粗放，燃料管控能力不强

火电企业 70%以上的成本属于燃料成本，因此燃料成为了火电企业主要影响因素。燃料在火电企业的生产、经营、管理等方面发挥了极其重要的作用，也是关键的影响因素，抓住了火电企业的燃料管理就抓住了火电企业的命脉。因此，燃料的管理创新成为了火电企业的永恒主题，也是电力行业共同研究的方向。目前大部分发电企业燃料管控能力不强,数据实时性不足,掺烧手段单一,成为了困扰火电企业精细化管理的瓶颈。

数据来源

当前，大唐集团大数据平台数据来源主要包含三大类：

1) 物联网数据

发电行业是设备密集型、技术密集型、资金密集型企业，发电设备各种各样的传感器在每时每刻每分每秒每毫秒微秒都会产生海量实时大数据。以典型的 2×600MW 燃煤火电机组为例，它拥有 6000 个设备和 65000 个部件，DCS 测点数平均达到 28000 个（不含脱硝等新上环保设施）。我们仅考虑生产实时数据（不包括图像等非结构化数据），扫描频率 2 秒/次考虑，2×600MW 机组的年数据容量的实时数据库数据容量为 114GB。再加上水电、风电机组产生的数据，保守估计大唐集团一年的生产实时数据超过 200TB。

2) 业务系统数据

业务系统数据包含 ERP 系统、综合统计系统、电量系统、燃料竞价采购平台相关的设备台账数据、发电量数据、燃料竞价采购数据等，每年 500G 左右。

3) 外部数据

外部数据包含地理信息数据、天气预报数据等，每年 500M 左右。

技术方案

大唐集团自主研发的大数据平台 X-BDP，基于 Hadoop 研发的企业级大数据可视化分析挖掘平台，是一套集数据采集、数据抽取、大数据存储、大数据分析、数据探索、大数据挖掘建模、运维监控于一体的大数据综合平台。



图 5 大唐集团工业大数据技术架构

平台应用大数据、云计算、物联网、人工智能等关键技术，提供多种存储方案和挖掘算法，支持结构化数据、半结构化数据和非结构

化海量数据的采集、存储、分析和挖掘，提供多种标准的开放接口，支持二次开发。平台采用可视化的操作方式，降低数据分析人员和最终用户使用难度。

1) 研发了智能数采通、智能隔离网闸等支撑设备，为大数据平台的搭建奠定了基础

开发了数据采集装置与系统、隔离网闸、实时数据库、数据甄别装置与算法、电厂机组性能数据分析装置等新产品，并得到大规模应用，大部分技术与设备获得国家专利。

为解决电力自动化系统中设备在通讯协议复杂多样化的情况下相互通讯、控制操作与通讯标准化的问题，研发了一种电厂数据采集装置、一种电厂数据采集系统和一种用于电厂的具备安全隔离功能的数据采集装置，保证数据稳定、可靠、实时得地进行数据采集。

为解决复杂网络环境数据安全传输系统，研发了复杂网络环境数据安全传输系统提供集数据采集，传输，保存等功能于一体的整套解决方案。

为对快速变化的实时数据进行长期高效的存储和检索，自主研发了实时\历史数据库（**X-DB** 实时数据库），支持百万千万级标签点，采用独有的 **X-BIT** 按位无损压缩编码算法，保证了数据不失真的情况下，提高数据的压缩率。

为能够主动地发现测点异常，研发了一种在线测量数据准确性甄别方法、一种改进 **53H** 算法的数据检验方法和一种数据甄别与预处理物理卡。

2) 通过互联网技术、传感器技术，对实时能源数据的汇集与分析，形成了大数据分析平台

通过互联网技术，应用智能数采通，实现对大唐集团国内外所有发电设备主要生产实时数据（含环保数据）的集中和统一，并以此为基础建立集团公司生产数据中心，实现生产数据的有效链接、集中、共享和应用。通过互联网技术、传感器技术，对实时能源数据的汇集与分析，把大量分散在生产现场的测量仪表作为网络节点，构建了覆盖全集团、全过程的生产、燃料、资金智能互联网体系，实现了人与发电机组、机组与机组的有机结合，提升了发电企业的效益。

3) 通过自适应模式识别算法，实现了机组远程诊断与优化运行

根据发电设备海量的历史运行数据，采用先进的模式识别算法，建立了数据驱动的自适应于该发电设备的运行状态模型。完善了包括 SIS 与辅控网在内的互联网智能功能，实现了电厂设备与互联网的信息融合，实现了电厂设备状态实时监视，捕捉设备早期异常征兆，实现设备运行健康度监测、劣化趋势跟踪及设备故障早期预警，以提高现场优化运维水平。基于智能挖掘的电力生产优化诊断技术，构建了发电机群远程集中故障诊断和设备优化运行的非线性状态估计等数学模型。

4) 通过智能进化多目标优化算法，实现了燃料“采制化存”和配煤掺烧最优方案

基于燃料的消耗情况及实时价格等数据，通过智能优化算法达到最优的燃料采购方案及储存方案，为发电企业节约资金。通过物联网

技术搭建的 GPS 平台技术，实时对运输公司的车辆路线监控，减少运输过程中电煤的损失。在满足配煤掺烧的约束条件下，系统使用智能进化算法对配煤掺烧进行多目标优化，自动生成同时满足单位质量燃料成本最低、锅炉燃烧效率最高以及氮排放量最少的配煤掺烧最优方案。

应用效益

1) 经济效益

大数据、工业互联网与发电工业融合创新显著提高了效率，使得生产力获得巨大提升。

应用智能进化算法对配煤掺烧进行多目标优化，自动生成同时满足单位质量燃料成本最低、锅炉燃烧效率最高以及氮排放量最少的配煤掺烧最优方案。2015 年掺烧经济煤种原煤量共计 6398.34 万吨，每吨节约 32.5 元，节约成本 20.81 亿元。2016 年掺烧经济煤种原煤量共计 8296 万吨，每吨节约 23.11 元，节约成本 19.34 亿元。

通过自适应模式识别算法的实时机组远程诊断优化技术，保证了机组的连续安全运行，提高了机组经济性。2015 年，大唐集团发电量 3793 亿千瓦时，入厂煤的平均单价 463.23 元/吨，发电煤耗同比下降 2.55g/kWh，节约 4.48 亿元。2016 年，大唐集团发电量 4699 亿千瓦时，入厂煤的平均单价 407.79 元/吨，发电煤耗同比下降 2.34g/kWh，节约 4.48 亿元。

2) 社会效益

通过对设备污染物排放数据的实时在线监控，实现排放超标的预警，同时由于采用了最优的配煤掺烧方案，且设备处于最优的运行状况，极大降低了污染物的排放。自应用以来，平均每年减少二氧化硫排放 84.98 万吨、氮氧化物排放 125.34 万吨。

由于发电企业的独特性，遍布全国各种各样的传感器在每时每刻每分每秒每毫秒微秒都会产生海量实时大数据，构建了一个庞大的能源互联网，中国大唐集团公司通过大数据应用，推动发电工业领域乃至整个工业领域的创新，加速能源市场自由化的进程，为中国发电工业升级转型提供了典型范例。

7.4 东方国信大数据实现联合利华能效提升

基本情况

东方国信是中国上市公司、大数据龙头企业，成立 19 年来一直坚持自主研发。其业务遍及全球，为全球 42 个国家和地区的 1000 多个客户提供大数据产品与服务。产品和服务包括 BI（商业智能）、CRM（客户关系管理）、能效管理软件、数据库软件、大数据解决方案等。主营业务覆盖通信、金融、智慧城市、公共安全、工业 4.0、医疗、大数据运营、新闻媒体和农业。

科托帕希公司是东方国信全资子公司，已为全球超过 35 个国家的工业客户提供了能效管理大数据服务，包括软硬件安装、调试和管理以减少能源消耗。能效管理软件 **Strata** 是基于先进的工业大数据、网络通讯、智能硬件、云存储等技术，实时采集工业现场的数字仪表、

传感器数据并抽取现场既有自控系统记录进行海量数据在线分析和挖掘的企业级平台，助力企业进行管理对标和优化，为企业能源管理及设备效率提升与废弃物减量化、生产计划安排等提供实时在线分析、决策平台。平台综合 CI、OEE 及能源专家系统模块构成生态系统。

联合利华集团采用科托帕希的大数据服务，在全球 100 多个工厂实施了能效管理系统建设，将各分厂的能效数据采集并传送到总部云平台数据库，对所有数据的分析处理都在此平台上进行。不仅仅是能耗数据的展示和监测，而是对能源的生产、转换和消耗进行在线平衡，从能源系统角度发现节能机会。

业务痛点

联合利华集团的生产技术和设备非常先进，能源使用效率也排在世界前列。为进一步降低生产成本、提高产品竞争力，联合利华提出每年在上一年基础上减少能耗 5%。在多年技术升级改造的基础上，能源的生产和使用已经成为一个庞大的系统，进一步的节能减排变得越来越困难。通过常规的技术改造已经没有节能潜力可挖，时有头痛医头、脚痛医脚的情况发生。

为了实现节能目标，联合利华采用科托帕希能效管理大数据分析系统，以客观数据为依据，以能量平衡为基础，从系统角度对海量数据的分析、比对，发现潜在的节能机会。对机会的风险和收益作为一个整体进行评估，并制定实施计划，通过大量的机会识别对节能目标做到了分解落实。

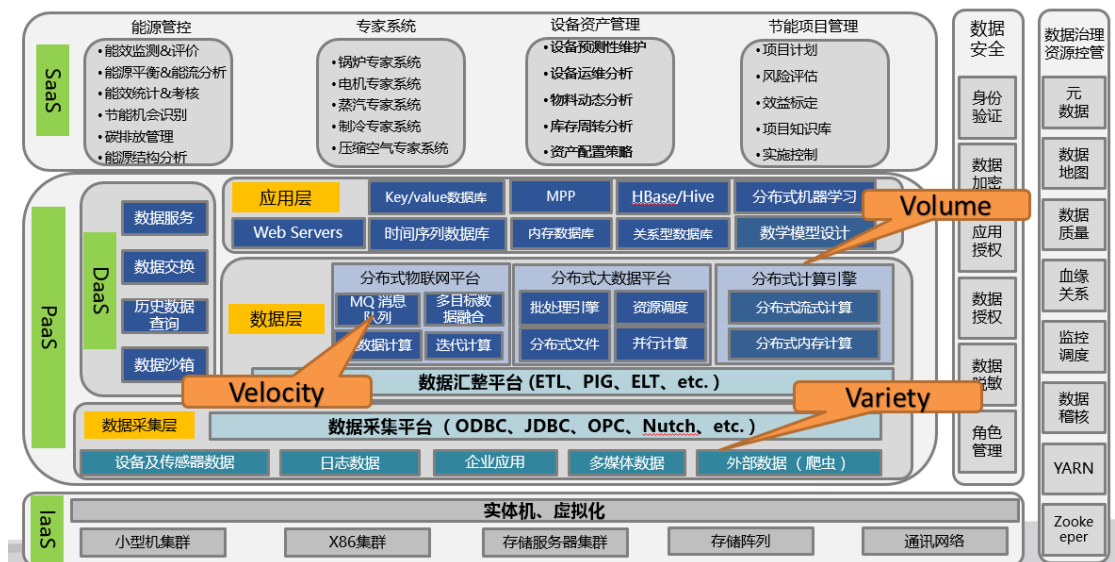
数据来源

数据主要来源有：(1) 机器数据：设备及传感器数据、PLC、DCS；

(2) 业务数据：BMS、MES、SCADA。

采集模块或采集软件定时采集仪表或控制系统能效数据，将全厂能效数据进行汇总，统一发送到云平台上的数据库。系统支持各种标准接口协议，从仪表或已有控制系统采集数据，也可以通过手工录入抄表记录，支持将 Excel 历史数据批量导入，将能效相关数据采集到数据中心。每个工厂平均 200 多个数据采集点，为了满足后台计算需要，更新频率采用 1 分钟一次，联合利华能效管理系统累计采集 14 亿数据点，1269 个活跃用户，实现全球工厂的能效对标。

技术方案



能效管理大数据系统包含：系统应用支撑、现场数据采集、因特网传输和云平台存储、数据应用计算分析、用户网页访问五部分。具

体应用模块包括：

1) 综合能耗分析与预测预警

Strata 可使用各种分析模型，识别工艺积极或者非积极生产状态，找到与能耗变化高度相关性的关键参数，比如产量、度日数等并建立合理的能效绩效目标来监控能源消费，同时可计算单位产品能耗，根据产品设置回归分析的目标值，并设置警报，对比能源消耗和产量发现节能机会，也同时可以实现聚类、分类、关联、例外、时间序列、空间解析等丰富的分析模型。在联合利华某洗发液工厂，通过产量和电耗的回归分析，预测基础负荷，设定理想电耗曲线，使该厂合理安排每天的生产负荷，仅此一项就节电 15%。

2) 机会识别、量化与节能量监测

对各类生产设备进行实时能耗监控。并根据总体消耗及分类消耗能源数据对比，识别能效改进方法，并对此做出量化。同时利用累计和图分析技术监测累计偏差，用以分析积极或者消极的能耗趋势，量化分析和监测节能量和浪费情况。在某食品厂识别出 18 个节能机会，通过调整生产运行方式以及技术改造，全年节约能源开支 50 万欧元。

3) 故障预测和设备整体效率分析

通过夹点，资产可用性，瀑布模型，计划、非计划停机深度分析报告打破原有的管理方法，使设备管理各个环节得到系统性提升。为联合利华节约大量的维修和停产费用。

4) 能源专家系统

基于大数据分析，通过不同类别能源数据计算，形成了两千多个

方案的知识库。在此基础上具备了锅炉专家系统、电机专家系统、蒸汽专家系统、制冷专家系统和压缩空气专家系统。

锅炉专家系统通过耗差分析实时调整运行参数，使原有的运行模式受到重新评估，监视蒸汽负荷状态进行优化，确保锅炉始终处于最佳运行状态。在实施的工厂中锅炉效率平均提升 2%。

电机专家系统基于系统负载因数和运行状态(运行小时数、相差、震动、电机温度、启停次数)，自动生成能耗及可靠性的改善建议。节能 3%-10%，可靠性提高 20%，停机减少时间 5%-7%，减少电机配件 30%，维护费用降低 20%。

蒸汽专家系统实时计算蒸汽平衡，对散热、泄漏等超标损失及时发现并报警。系统疏水器实现在线监测，彻底杜绝了因疏水器失效造成的蒸汽损失。对蒸汽系统通过建模做到了逐级使用，闪蒸汽、凝结水回收率达到 95% 以上。实施后，蒸汽系统整体效率提高了 10%。

制冷专家系统提供了一个从全面的生命周期角度出发的具有可靠性的制冷系统。包括了整个制冷系统中每一个阶段的能量转换过程。通过监控制冷环境并优化的冷却负荷，工厂实现可靠和高效方面的改进。将 PLC 系统与 Strata 相连接，通过 HMI 和传感器阵列计算出实时的系统性能系数，优化冷凝和蒸发温度来减少末端冷负荷。通过改进关断、隔离和机械的序列进一步提高效率。联合利华某冰激凌厂采用制冷专家系统，第一阶段通过实施新的运行组合方式，在 2016 年全厂节电 6.2%，相当于每年节约 6 万欧元。第二阶段通过优化滑阀控制策略实现节电 14%，相当于全年节电 21.7 万欧元。

项目技术先进性可总结为以下三点：

1) 结合云计算、大数据技术，通过数据挖掘和数据分析，以数据为依据，寻找节能机会。

2) 使用云平台，除了企业内部使用，还会邀请行业专家对工厂能效使用情况进行分析，提供解决方案。

3) 应用 **Strata** 平台作为分析工具和辅助决策工具。该平台有更好的交互性、易操作性（通过简单拖拉拽实现），能无编码操作，界面友好，支持个性化定制，用户体验好。

应用效果

截至目前，已经完成联合利华全球 100 多家工厂的能效数据接入和分析。已接入的碳排放数据占联合利华全部排放的 50%；已接入的能耗数据占全部能源消耗的 54%；已接入的水耗数据占全部水耗的 64%。大大提高联合利华的能源管理效率和效果。平均为每个工厂实现能源节约 5%-15%，节水 5%-30%，原材料减少 1%-3%，包装节约 5%，取得了巨大的经济效益和社会效益。

7.5 沙钢能源管理大数据应用

基本情况

冶金自动化研究设计院（以下简称“自动化院”）创建于 1973 年，是工业自动化领域大型研发机构，拥有国家重点实验室、国家工程实验室、国家工程技术研究中心科研平台。其大功率交流变频调速、智

能控制、现场总线、信息化、余热余能回收利用、交流伺服装置、冶金专用仪表、大型液压伺服缸及伺服系统、弹性胶体等技术在国内具有领先优势，相关工程和产品广泛应用于冶金、机械、电力、建材、能源、军工等行业，遍及全国 29 个省、市、自治区。

江苏沙钢集团有限公司（以下简称“沙钢”）是目前国内最大的电炉钢和优特钢生产基地，也是江苏省重点企业集团、国家特大型工业企业。主导产品“沙钢”牌宽厚板、热轧卷板、高速线材、大盘卷线材、带肋钢筋等已形成数十个系列和 300 多个品种。目前，公司拥有总资产 1050 亿元，职工 17000 余名，年产铁、钢、材的能力分别为 1810 万吨、2160 万吨和 2225 万吨。

早在 2006 年，自动化院就为沙钢建设了能源管理中心，实现了企业能源生产、输配和消耗的动态监控和管理。2013 年起，沙钢投资上亿元逐步完善了能源计量仪表。在此基础上，2015 年自动化院为沙钢全面升级了能源环保管理系统，建立了能源环保大数据平台，普遍提升了数据分析挖掘和应用水平，使沙钢获得了可观的经济和社会效益，为钢铁企业能源中心的升级指明了方向。

业务痛点

由于行业产能过剩、竞争日趋激烈，钢铁企业的成本压力越来越大；应对气候变化、治理大气污染，钢铁企业的节能减排压力也越来越大。主要面临如下问题和机遇：

- 1) 工艺节能、设备节能潜力下降，投资收益率降低

钢铁企业陆续进行了设备大型化、现代化改造，普遍采用了变频节电、TRT、干熄焦发电、余热发电、高炉大喷煤等技术，通过技术改造实现节能降耗的潜力越来越小，投资收益率越来越低。

2) 能源管理较粗放，科学化、定量化过程管理能力不足

长期以来，钢铁企业形成了生产为主、能源为辅的管理理念。能源管理较为粗放，表现为：能效指标体系缺乏系统支撑，对节能潜力点挖掘不足，数据不真实掩盖了问题；节能管理搞运动、一阵风，缺乏信息系统的监督、科学评价，节能管理的成果难以持久；能源信息分散、共享程度不足，使得公司能源管理、分厂能源管理各行其道，未形成上下贯通、全员参与的局面；能源调度决策缺乏模型支持，调度人员难以进行预先的、精确的调整。

3) 管理节能潜力巨大

钢铁企业设计时大马拉小车问题突出，能源供应超出实际需求，检修期间浪费更加严重，降低介质单耗潜力大；大工业用电按照峰平谷分时计价，峰电价一般为谷电价 3 倍，避峰就谷生产的经济效益十分显著；蒸汽管网保温不合理、阻力损失大、跑冒滴漏等造成管损比例大；一边外购天然气，一边放散煤气，一边外购技术气体，一边开着低效的空压机组等等。

在严峻的形势面前，钢铁企业唯有利用大数据技术提升能源管理科学化水平，降低能耗和能源成本才能立于不败之地。

数据来源

沙钢能源管理环保数据有三大类：

1) 计量仪器仪表、环保数采仪等机械数据

能源计量仪表共约 1500 块，数据更新频率每秒一次。其中，含天然气、煤气、氧氮氩、压缩空气、蒸汽、水等非电介质计量仪表约 900 块：主要采集压力、瞬时流量、累积流量。电表约 600 块：主要采集有功功率、峰平谷总有功电量。环保数采仪约 25 套，数据更新频率每分钟一次，其中烟尘数采仪主要采集 SO₂、NO_x、O₂、烟尘、温度、压力、流量，废水数采仪主要采集 PH、COD、TOC、氨氮、总磷、流量、累积流量。

2) 检化验、产量等业务数据

检化验数据包括煤气化验成分、热值，氧氮氩纯度、水质等，数据更新频率每班一次；产量数据包括所有车间每日产品及副产品产量，数据更新频率每日一次。

3) 指标体系计算数据

在原始数据的基础上，系统需要计算各种能源环保指标数据，包括峰平谷用电比、介质平衡率、介质单耗、单位能耗、总能耗、综合能耗、设备峰平谷总运行率、排放达标率、计划命中率等，这些数据归口粒度分为分厂、车间、重点设备三级，计算频度分为每班、每日、每月，计算数据量约为采集数据量 3 倍。

实时监控、日常管理、财务核算对数据的频度要求不同，数据在存档时有不同时间粒度（秒、分、时、班、日、月）等、不同类型（原始值、差值、修正值、平衡值等）的多个存档点。截至 2016 年 12 月，

系统标签数量 7300 个，存档点数 35000 个，秒级数据压缩后保存 1 年，环保数据保存 3 年，其它数据保存 10 年以上，系统的总数据量在 200 亿条以上。

技术方案

沙钢能源管理大数据应用以实时数据与业务数据融合为基础，集实时数据库与关系数据库功能于一身，打造数据平台。进而提供了丰富的业务分析与数据应用功能，特别是节能管理专项工具集、多介质预测调度模型，大幅提升了沙钢能源管理水平。

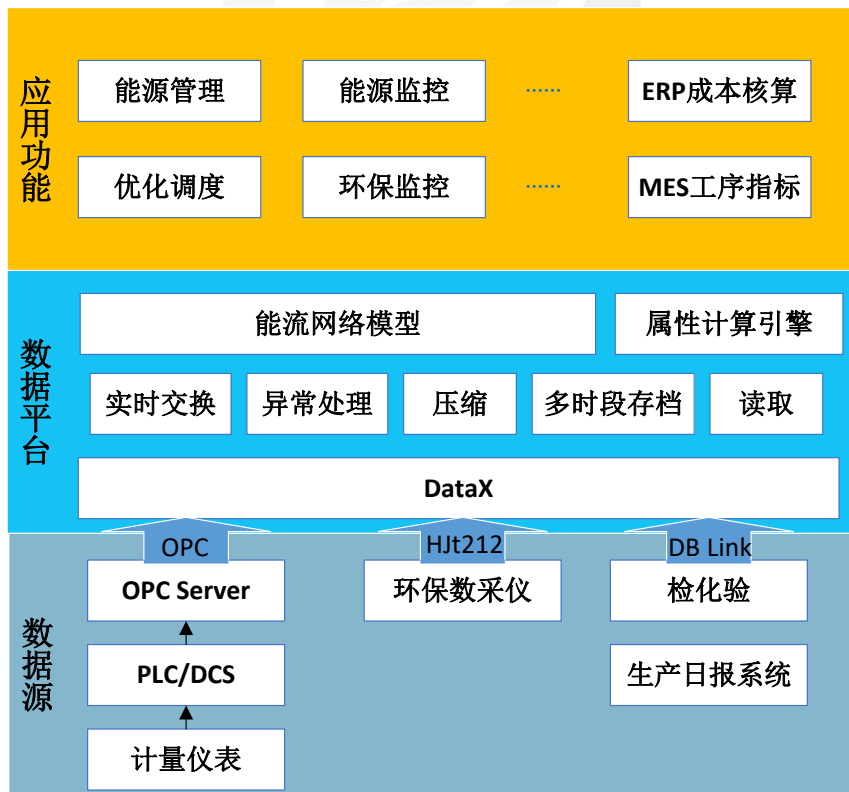


图 7 沙钢能源管理大数据技术架构

1) 实时数据与业务数据高效一体化管理

自动化院开发的能源大数据平台软件，实现了实时数据高效采集、

异常过滤、压缩、多时间尺度存档；采用能流网络模型对数据进行再组织，使能源管理应用有了统一的语义环境；动态属性计算技术使指标能够在查询时由计量数据不落计算出来，根本上保证了数据真实性；针对通讯故障造成的数据错误设计了自动修复机制，代替了人为修复工作，保障了数据质量。

2) 钢铁企业节能管理特色工具集

针对余热余能回收与发电，提供了转炉煤气回收跟踪管理、发电机组效率跟踪管理功能；针对避峰生产，提供了车间峰平谷用电统计分析、煤气柜移谷填峰发电跟踪管理、设备（如原料输送皮带）峰平谷总运行率管理、车间避峰补水管理等功能；针对介质损耗，提供了每班计量平衡率统计分析功能，帮助发现跑冒滴漏以及仪表故障；针对过度供能，提供了车间、重点设备介质单耗统计分析、对标管理。

3) 多介质预测调度技术，提升能源调度智能化水平

自动化院研发了工况组合预测技术、多介质协调调度技术，实现了煤气-蒸汽-电多介质协调调度，可以根据生产计划、设备检修及故障信息生成调度方案，指导调度人员进行预先、精确的调度调整，降低了管网压力波动，降低煤气放散；优化发电机组煤气分配，同等条件下多发电；安排煤气柜谷段充气、峰段发电。

大数据平台采用内容为王的理念，突出数据弱化功能模块，为公司、分厂、车间、职能部门提供了 500 余个分析表格，为所有数据点提供了趋势曲线，简单实用，深受用户青睐。

应用效益

1) 经济效益

2016年，沙钢通过大数据应用提升能源管理水平，获得了3.49亿元经济效益。

①吨钢转炉煤气回收增加 20m³，经济效益 9945 万元

通过系统对四个转炉炼钢车间吨钢转气回收指标进行每日管控，吨钢转气回收率从 112m³/t 提高到 132m³/t，热值从 7200kJ/m³ 下降到 6950 kJ/m³。按照 1600 万吨转炉炼钢量计算，相当于多回收热量 177.6 万 GJ。按照自发电效率 12500kJ/kWh、外购电 0.7 元/kWh 计算，此项经济效益为 9945 万元；

②峰电比降低 2.3%，经济效益为 9660 万元

通过系统对电炉轧钢冲击负荷错峰用电、煤气柜移谷发电、原料输送皮带错峰运行、车间错峰补水等削峰填谷专项进行管理，使公司总的峰电比下降 2.3% (从 33.3% 下降到 31%)，谷电比相应上升 2.3%，峰谷电价差 0.6 元/kWh，年外购电量 70 亿 kWh，此项经济效益为 9660 万元；

③蒸汽管损下降了 50t/h，经济效益为 5256 万元

通过系统随时跟踪蒸汽系统计量平衡率，发现跑冒滴漏后更换了 500 个疏水阀，增加一段联接管，使蒸汽管损下降了 50t/h，按照每吨 120 元计算，经济效益为 5256 万元；

④少开 5 台制氧机组，经济效益为 10000 万元

通过系统跟踪氮气单耗和压力变化，在满足工艺需求的前提下降低管网压力，停开了 5 台 20000m³/h 制氧机组，每台机组节约 2000 万元用电成本，此项经济效益为 10000 万元。

2) 社会效益

通过大数据应用，2016 年沙钢综合能耗水平从吨钢 580 公斤标煤下降到吨钢 576 公斤标煤，节约能源 8.4 万吨标准煤，节能减排效果非常显著。

7.6 陕鼓动力智能运维大数据应用

基本情况

西安陕鼓动力股份有限公司属于陕西鼓风机(集团)有限公司的控股公司，公司成立于 1999 年，2010 年挂牌上市。是为石油、化工、冶金、空分、电力、城建、环保、制药和国防等国民经济支柱产业提供透平机械系统问题解决方案及系统服务的制造商、集成商和服务商，属于国内透平行业领军企业。本案例以陕鼓动力在徐州东南气体有限公司投运的智能远程运维系统为例，介绍作为动力设备制造厂商依托设备大数据系统平台建设向设备智能服务商成功转型的具体措施。

徐州东南气体有限公司隶属于秦风气体股份有限公司，是为徐州东南钢铁工业有限公司 2x1080m³ 高炉配套建设的专业制氧企业。装置规模为 2x2 万 Nm³/h，主要产品为氧、氮、氩气体和液体，一期核心动设备为一套空压机组与氮压机组。

自 2003 年以来，陕鼓动力始终致力于远程监测、故障诊断、网络

化服务平台、云服务应用需求等方面的技术与相应技术的推广应用，主要经过三大重要阶段：系统推广应用期（2003年-2010年）、网络化诊断与服务平台建设期（2011年-2013年）、设备全生命周期大数据建设期（2014年-至今）。2013年已成功解决了面向动力装备的多地多通道并行信号接入、现场海量数据的采集与压缩存储、远程传输、远程升级、智能故障诊断算法与基础应用、数字化检修维修、备件预测协同信息化管理等技术的研发与应用。

业务痛点

徐州东南气体有限公司的动设备管理始终是影响气体生产的核心困扰难题，设备意外停机一天，会造成生产企业 1000 万元以上的直接损失。因缺乏专业的设备管理与维修人员，多次意外设备故障造成生产损失赔偿金额高达 400 万元以上。

1) “定期检修”传统设备维护模式易造成维修成本过高，工期无法控制

定期维护模式下，常出现“过度维修”与“维修不及时”的情况，两者均会直接影响设备有效运行时间，事后维修会造成工期延长，均导致运行维护成本的上升，严重影响用户生产主业。

2) 缺乏健全的设备全生命周期管理档案，维护信息碎片化，断序严重

由于缺乏专业化检修人员，大多数空分企业设备检修通常采用外包方式进行，在没有设备大数据全生命周期管理系统支持的情况下，

容易造成设备维护管理信息不连贯，没有继承性，碎片化严重。

3) 传统设备维修外包服务，容易丧失企业对设备资产的掌控

没有设备信息数据系统支持的设备外包通常会隔绝用户对设备状态的感知与把握，使用户对外包方产生依赖，不断削弱对设备维护成本的掌控。

综上所述，空分行业的设备故障发生率，占整个工艺装置故障的92%，设备管理对空分企业来说是各项业务的重中之重，徐州东南气体有限公司的设备管理问题在空分行业普遍存在，属于典型的行业难题。

数据来源

陕鼓智能远程运维系统主要包含三大类数据：设备状态大数据、业务数据与知识型数据：

1) 设备状态数据

针对于动力设备的快变量数据，利用陕鼓自主研发的 IMO1000 系统进行高通量数据的采集，采集信号主要为振动传感器电压变化值，采集速率每振动测点达到 10K/S。慢变量数据主要指设备与装置的工艺量与过程量，利用 TCS 装置级数据采集系统，通过从机组的 DCS 系统获取，刷新频率在 1 秒到 3 秒。一套空分装置，测量数据点共计 336 个，实时原始数据量达到 10M/S，设备启停机或故障时，最高数据通量峰值在 35M/S 以上。

2) 业务数据

业务数据主要包括用户档案、机组档案、现场服务记录、用户合同管理、备件生产管理等设备管理过程中产生的数据，主要来自陕鼓工业服务支持中心的客户管理与服务管理系统。

3) 知识型数据

知识库数据主要包括设备设计图纸、加工工艺、装备工艺、制造质量数据、测试数据、核心部件试车、整机试车、各类标准工时文件等。该部分数据以 IETM 系统管理为主，以 PLM、CAPP、ERP 数据为补充。

业务数据与知识型数据总量约为 1TB，且更新随业务流程状态改变，每日平均增量在 5M 以内。

技术方案

为了确保用户机组的安全稳定运行，避免由于网络不稳定造成关键实时预警的漏报误报，本系统平台采用原始数据本地存储，处理，预警；关键数据实时同步压缩上传的接入模式。即使由于网络问题造成通讯中断，现场系统仍然可保证实时进行分析预警，对突发的故障数据进行记录与处理，确保用户机组的万无一失。该设备大数据平台技术框架图如下。

如图所示，通过设备振动、温度、流量、压力等传感器与控制系统，将数据接入到 IPMC 系统，数据实时处理后，送入现场监控一体化 HMI 系统，可直接向用户呈现设备运行状态分析结果。同时，利用互联网或 3G/4G 无线网络，将数据实时远传至陕鼓远程智能运维

中心，中心专家结合 IETM、备件协同系统、PLM 等其他数据，向用户提供中长周期的设备运行指导意见。

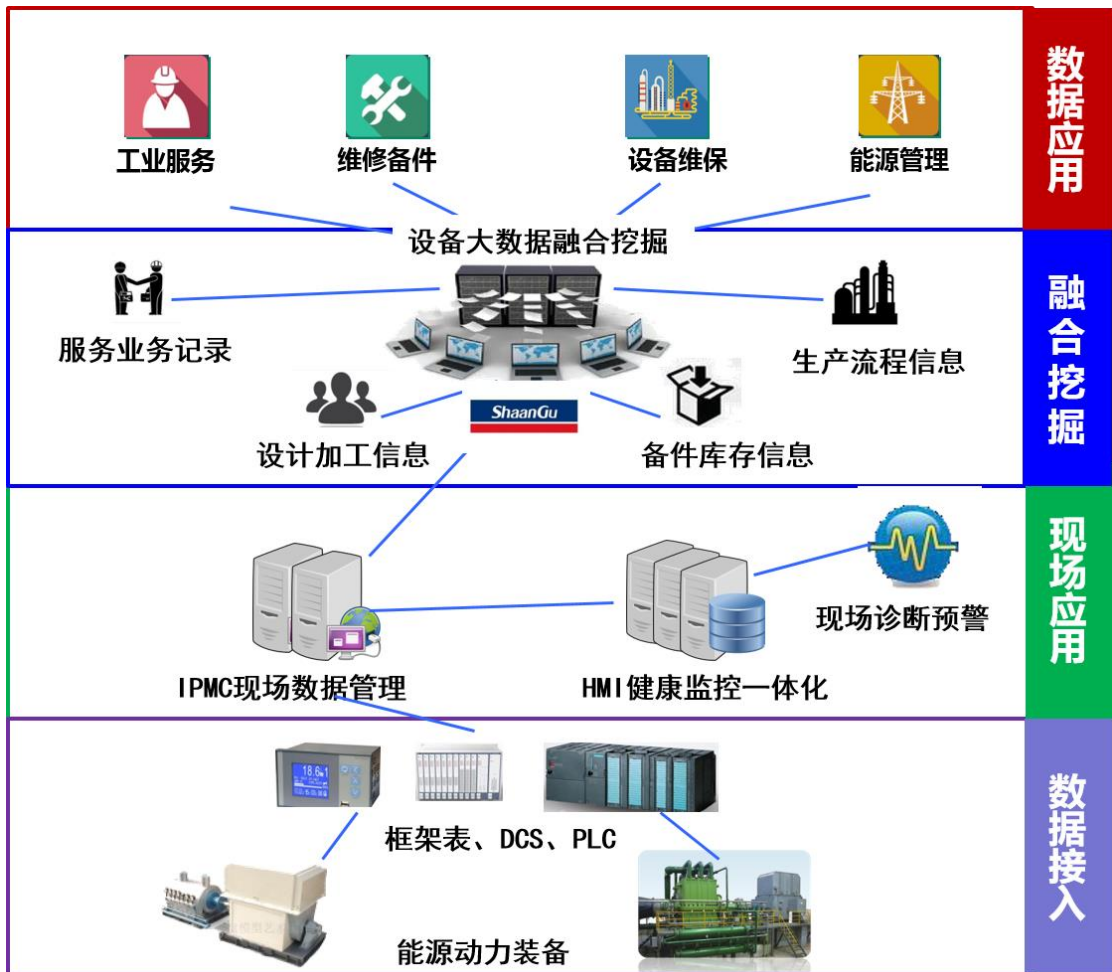


图 8 陕鼓远程智能运维系统技术架构

本项目产品具有以下三个方面优势特点：

1) 高通量实时数据采集处理

由于动力装备的转子动力学状态分析要求，需要对转子各测点振动数据进行高速并行同步整周期 A/D 采集及滤波调制等处理。对实时数据清洗、有效性判别、自适应调整预警门限等技术要求较高。

2) 成熟的大数据挖掘应用

动力装备领域数据类型多、数据差异化大，建模相对困难，且需

结合转子结构模型、动力学模型完成复杂数据关联性分析、故障根原因分析等大数据挖掘分析应用，才能得出对检维修具有指导作用的准确方案。

3) 多信息融合健康故障诊断与设备性能优化

动力装备领域故障诊断时需用到大量专业的图谱工具，数据分析相对困难。同时，因工艺的复杂性，需进行与工艺量关联分析、自适应预警门限、性能对标、喘振预警与优化控制、性能仿真等。

应用效益

自 2012 年底首次建立试点以来，陕鼓已在 13 家用户得到推广验证，其中包括秦风气体的八家空分企业、宝钢湛江、山西平朔煤炭等冶金、化工企业集团，四年来为陕鼓产生维保服务收益四千余万，带动备件、检维修等业务收益三亿三千万，为用户带来直接设备收益约一亿七千余万。

以 2016 年徐州东南气体有限公司服务数据测算：

用户直接设备收益包含两部分：设备维护费用的降低与设备对生产的影响收益。

1) 节约设备维护成本：2016 年度两套机组年度设备大修费用与备件费用维护成本为 144 万元（按市场价格测算）。采用陕鼓远程智能运维服务，一年维护费用为 72 万元，节约支出 50%。

2) 增加业务收益：减少非计划停机 50% 以上，延长设备有效运行时间 15 天，增加业务收益 74 万元。另外，采用预知性维修，精确制

定维修计划缩短正常检修工期 51.6%，带来生产效益 77 万元。

综上，本系统 2016 年共为用户一套空分装置增收产值 223 万元以上，另外还为用户节约管理内耗 47%，节约设备管理人力成本 60%，减少生产保险投入 15%。

该运维服务平台属于利用大数据支持智能服务经济的转型探索，对重大装备制造型企业向智能服务转型起到良好的示范作用。主要意义包括：在提升智能服务技术与积累核心服务技术经验方面依托设备大数据分析支持，可以提升远程诊断服务、检维修服务及备件零库存服务中的工作效率、降低服务成本，并通过对设备大数据的积累与挖掘，不断提升装备制造企业的核心竞争力，带动产业结构优化升级；在推动动力装备智能化进程方面：安装有智能信号采集与监测诊断系统的动力装备，结合制造厂商远程诊断中心大数据挖掘、智能商业应用软件及远程诊断服务支持，可以为装备提供一定的预测、感知、分析、推理、决策功能，促进传统的产品服务逐渐向产品运行管理、决策分析和优化产品设计等方向渗透，最终加快装备智能化的进程。

7.7 仪电显示工业大数据应用实践

基本情况

上海仪电显示材料有限公司成立于 2007 年，是中国大陆首家也是唯一一家五代线液晶显示面板配套彩色滤光片独立生产厂商。公司主导产品为五代(尺寸为 1100*1300m²)液晶显示面板彩色化核心部件-彩色滤光片，覆盖了智能手机、娱乐影音、平板电脑、车载显示、桌

面显示、工控显示等应用领域，已成为我国 TFT-LCD 关键原材料国产化战略布局中的关键力量，国内市场占有率达到 35% 以上。

上海仪电集团已经制订了智能制造战略，并打造智能制造板块。仪电显示将在智能制造领域不断尝试，并形成可推广可复制的运营模式。公司智能化发展分为三个阶段，从自动化，到信息化，最终发展为智能化。目前正在进行第三阶段的智能化转型升级。

公司工业互联网起步较早，经过多年打造，公司信息化与工业系统已进行了深度融合，为数据价值挖掘和智能制造的转型升级奠定了良好的基础。2007 年仪电显示建立了贯通 23 条生产线的自动化生产；2011 年开始构建“数字+精益”的工业大数据平台，自主研发了数字化精益管理方法和先进的工艺控制系统，实现了客制化订单小规模柔性自动化生产；2016 年开始进行智能化升级，引入了无线物联网，建设了虚拟工厂，并搭建了新一代的大数据平台，将全面联通生产运营系统、能源系统、物联网系统和虚拟工厂，在工业大数据的深度应用进行着探索和实践。

业务痛点

电子产品最近几年发展迭代速度非常快，呈现出小批量多批次，个性化和专业化的发展趋势，为仪电显示机械化大规模生产带来机遇和挑战：

- 1) 需要提升机械化生产的柔性以满足个性化订单

公司所在的行业是液晶面板行业，这个行业最大的特点是信息化

和自动化程度比较高，整个行业面临着在全自动化的生产方式情况下，迎接个性化和专业化的市场发展趋势。仪电显示通过工业大数据技术，对设备、物料、人员作业、环境、运营等数据进行深度挖掘，在信息技术与工业系统深度融合的基础上，实现以数字为驱动的运营模式，实现了接单零库存小批量多批次的自动化生产。

2) 增强在线工艺控制能力，在满足个性化订单的同时提升良品率

自动化生产更适合工艺稳定的大批量生产，因为自动化生产速度快，由于每个产品的工艺路线和工艺条件不同，小批量多批次的个性化生产，在频繁切换后，如果工艺没有控制好，将造成一批次产品的不良。在实践中不断摸索，基于数据挖掘技术自行研发了工艺控制系统，对影响品质的工艺参数进行趋势管理，在没有发生问题时提前预警和纠偏。在满足个性化订单的同时，产品的良品率大于 95%。

3) 增强数字化精益管理，在满足个性化订单的同时提升设备稼动率

个性化的订单带来的是小批量多品种的生产，以及大量的产品切换，每次切换都意味着产能损失。通过实践发现繁杂化—简单化—流程化—一定量化—信息化的精益管理方法在自动化生产产能提升方面效果并不明显。通过实践摸索，自创了数字化精益管理方法，将管理活动最小化，最小化的活动数字化，数字化的活动参数化，参数化的活动控制化。通过这套方法的运行，目前公司每月切换高达 80 多次，自动化生产的设备稼动率接近 80%，生产能力已经突破了设计产能

25%。

数据来源

当前，仪电显示大数据平台数据来源主要包含三大类：

1) 信息系统数据

主要包含了从设计到产品交付环节数据，主要有 ERP、CIMS、SPC 等。系统已累积近 10 年数据，在线保存 6 个月 3TB，存量数据约 40TB，数据毫秒级更新。

2) 能源系统数据

在建厂之初只有总厂端的能源消耗，2016 年选取了一条线做试点，应用无线物联网采集设备端进行水、电、汽的消耗，数据量已累积近 1 年数据，存量数据约 1TB，数据实时采集更新。

3) 物联网数据

主要包含实时采集的设备运行、生产制造、工艺品质、能源消耗、物料运输等信息。仪电显示的大数据平台已累积了近 10 年数据，在线保存 6 个月 3TB，存量数据约 40TB，数据通过内部光纤网传输，毫秒级更新。

技术方案

仪电显示“数字+精益”的工业大数据平台，从数据的全面性、准确性、多维度对数据进行深入挖掘和提炼，从管理和资源两个层级上进行：管理层级轴，通过大数据可视化系统进行及时的决策和创新；资源层级轴，通过大数据对设备、零部件、工艺等进行预测预防，确

保资源的充足供给，以达到数据驱动生产和管理创新的模式。



图9 仪电显示工业大数据平台架构

1) 通过大数据可视化平台，各层级管理人员能够及时发现随机或异常信息，进行及时决策和不断改进

通过统一的大数据可视化系统，各个层级管理人员能够快速直观了解到生产现场的全部信息，包括：设备、管道、消防设施的布局与运行状态；当前的生产数据，如当前执行订单、设备的工艺参数、产品质量情况、故障等异常信息、能耗数据、环境监测数据等；以及当日的统计数据，如设备综合稼动率、良品率、负荷度、综合效率等。一旦出现异常，及时采取相应措施，确保生产顺利进行。同时从中也能够发现需要不断革新改善的项目，达到旧价值改善，新价值开发的创新管理模式。

2) 设备运行状态、工艺参数、故障数据实时采集，基于大数据

分析，逐步实现设备预防性维护与产品良率稳定提升，确保资源充足供给

仪电显示基于预防失效模式与影响分析的理论方法和大数据技术建立了设备故障预警模型和设备远程维护系统，对设备和零部件进行预测预防，达到设备利用最大化；通过对质量和工艺进行预测和纠偏，提升在线质量监测能力，稳定并提升良品率。

工业大数据平台整体采用成熟的 Hadoop 分布式架构进行搭建。通过大数据引擎，对数据进行流式处理，满足数据分析的时效性要求；通过分布式运算架构，满足对海量数据在线和离线的深度挖掘分析。

3) 基于工厂仿真平台建立生产系统仿真模型，通过大数据分析，持续优化生产过程

目前，仪电显示基于数据自动采集系统和工业大数据平台，已积累了海量的生产数据，并利用工厂仿真平台建立了整个生产系统的仿真模型。该模型完全通过数据驱动与模型算法，实现了与生产现场完全一致的生产制程与生产调度逻辑，可以真实模拟实际的生产运行，对生产过程进行持续改善。

应用效益

“数字+精益”工业大数据平台的实践应用效果显著，为仪电显示带来了可观的效益：

1) 经济效益

① 生产柔性提升：机械化生产柔性提升了 8 倍，生产线产品切

换能力由原来的 10 次/月上升到现在的 80 次/月，实现了接单小批量多批次零库存生产。

- ② 技术质量、产量提升：能够满足市场更高分辨率液晶面板产品质量要求，产品分辨率从原来的 300 ppi 上升到 400 ppi。生产能力从原来的 550K/年上升到现在的 1001K/年，企业经济效益增长约 6000 万元/年。

2) 社会效益

离散型生产是机械自动化生产的难题，仪电显示“数字+精益”工业大数据的应用实践，实现了接单小批量多批次零库存的自动化生产模式，将为同行业和其他行业传统制造业转型升级提供参考。

7.8 中联重科工业大数据应用实践

基本情况

中联重科股份有限公司创立于 1992 年，主要从事工程机械、环境产业、农业机械等高新技术装备的研发制造，主导产品覆盖 10 大类别、73 个产品系列、1000 多个品种，工程机械、环卫机械均位居国内第一，农业机械位居国内前三。公司先后实现深港两地上市，是业内首家 A+H 股上市企业，注册资本 76.64 亿元。中联重科成立 20 多年来，先后并购了意大利 CIFA、德国 M-TEC、荷兰 RAXTAR、意大利 LADURNER 以及浦沅集团、新黄工、奇瑞重工等一大批知名的国内外企业，布局全球市场，打造全球化高端装备制造企业。

当前公司正积极推进“2+2+4”战略转型，立足产品和资本两个市场，

推进制造业与互联网、产业和金融两个融合，做强工程机械、环境产业、农业机械、金融服务四个板块，实现公司“产品在网上、数据在云上、服务在掌上”的新商业模式转型升级。

国内工程机械行业中，中联重科物联网起步较早，相关信息化基础较为完善，多年数据的持续积累为后续大数据分析和应用提供了良好基础。2006 年中联重科开始工程机械设备远程监控管理研究及应用，并建立物联网服务平台；2012 年实现业务系统运营环节全覆盖；2015 年建立新一代大数据平台，全面打通物联网平台、企业业务系统相关数据，并在此基础上实现了一系列深层次数据分析和应用，智能化应用水平处于国内行业领先地位。

业务痛点

工程机械市场进入转型期以来，行业面临如下问题和机遇：

1) 新机需求疲弱与巨大的服务后市场潜力

国内经济增速下滑，投资放缓及房地产市场疲弱，导致工程机械行业设备供过于求，新机需求放缓。而与此对应的是潜力巨大的服务后市场，按全国工程机械保有量 700 万台，维保费用 2 万元/台/年估算，市场规模近 1500 亿元/年。通过工业大数据技术，对设备、客户等数据进行深度挖掘，实现上下游信息充分共享和深度融合，降低成本的同时形成良性服务生态圈，将进一步拓展行业盈利空间。

2) 施工行业向规模化、集约化、专业化方向发展，对施工安全、效率、成本管控的重视程度不断提高

行业的调整将一批实力较差的“散客”淘汰出市场，市场集中度不断提高，专业化大客户比例增加。下游客户对施工安全、效率及成本管控的重视，要求设备厂商持续提升设备质量的同时，进一步强化设备智能化水平和数据分析处理能力，将服务从“被动服务”向“主动服务”升级，降低施工风险，提升无故障工作时间，实现“降本增效”。

3) 严峻的工程机械市场环境要求企业进一步精细化管理、高效科学决策，加速从传统生产制造型向高端服务型的转型升级

近年来工程机械市场需求持续下滑，以靠主机销售收入支撑发展的我国工程机械行业面临着严峻的生存考验。与此同时，我国工程机械行业客户也面临着工程项目减少，设备开工率与设备运行效率双低、运营管理成本及维修保养成本双高等困难局面，不利于行业客户的生存和发展。

严峻的工程机械市场环境对企业经营管理及决策提出了更高要求，如何通过大数据分析使企业更加贴近市场、更加理解客户，提升企业运营管理和决策效率，快速从传统生产制造型向高端智能服务型的转型升级是行业内每个企业面临的重大问题和挑战。

数据来源

当前，中联重科大数据平台数据来源主要包含三大类：

1) 物联网数据

主要包含中联重科设备实时回传的工况、位置信息。当前中联重科物联网平台已累积了近 10 年数据，监控设备数 12 余万台套，存量

数据量 40TB，每月新增数据 300GB。数据通过移动网络以加密报文方式回传，通过平台解析后实时保存至大数据平台。目前，数据采集频率 5 分钟一次，根据数据分析需要可进行调整，设备传感数据采集点将近 500 个。

2) 内部核心业务系统数据

包含了中联重科在营运过程中产生的业务信息，主要包含 ERP、CRM、PLM、MES、金融服务系统等数据，涵盖研发、生产、销售、服务全环节。当前，业务系统已累积近 10 年数据，存量数据约 10TB，数据每天进行更新。

3) 外部应用平台数据

包含了中联重科相关应用平台（官方网站、微信公众号/企业号、中联商城、中联 e 家系列移动 APP、智慧商砼、塔式起重机全生命周期管理平台）积累的数据、从第三方购买和交换的数据以及通过爬虫程序在网络上搜集的舆情及相关企业公开数据。除结构化数据外，平台还以日志方式保存了大量的用户行为数据。由于相关平台多于 2016 年推出，存量数据约为 1TB。

技术方案

中联重科工业大数据应用从“硬、软”两方面同时着手：硬的方面，通过研发新一代 4.0 产品和智能网关，进一步提升设备的智能化水平，丰富设备数据采集维度，提升设备数据采集和预处理能力；软的方面，基于大数据分析挖掘技术，形成多层次智能化应用体系，为企业、上

下游产业链、宏观层面提供高附加值服务。

1) 通过研发新一代 4.0 产品和智能网关,大幅提升设备智能化水平,夯实工业大数据应用基础

中联重科“产品 4.0”是以“模块化平台+智能化产品”为核心,深度融合传感、互联等现代技术,研发性能卓越、作业可靠、使用环保、管控高效的智能化产品。

通过在设备上加装大量的高精度传感器,实时采集设备的运动特征、健康指标、环境特征等相关数据,结合智能网关的本地分析功能,真正实现设备的“自诊断、自适应、自调整”。相关数据通过多种传输渠道(移动网络、WiFi、蓝牙等)回传至中联重科大数据平台,通过相关建模分析为客户提供包括设备实时定位、工况监控、油耗分析、设备异常分析、故障预警、工作运营统计在内的高附加值信息服务。

2) 打通多方数据,形成统一的工业大数据分析平台,对内辅助科学决策,对外提供智能化服务

中联重科大数据分析平台融合了物联网平台、业务系统、应用系统及第三方数据。分析角度涉及产品、经营、客户、宏观行业等方面,服务涵盖轻量级应用(中联 e 管家、服务 e 通等)和重量级专业领域应用(智慧商砼、建筑起重机全生命周期管理平台),并通过移动端 APP、PC 端、大屏幕等多种方式提供高效增值服务。

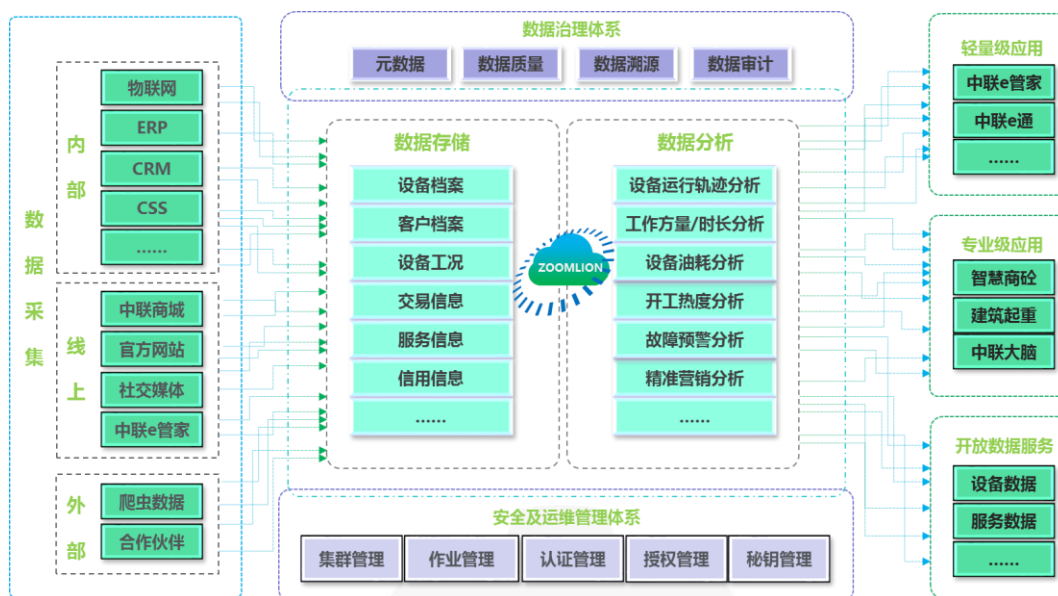


图 10 中联重科工业大数据平台架构

工业大数据平台整体采用成熟的 Hadoop 分布式架构进行搭建。通过流式处理架构，满足高时效性的数据分析需求；通过分布式运算架构，满足对海量数据的离线深度挖掘。前端通过统一接口层以多种通用格式对外提供数据分析服务。

考虑到大数据平台汇集了企业内外部多方敏感数据，为保证数据安全，平台引入了企业级数据治理组件，实现统一的元数据管理、数据质量控制、数据溯源、数据操作权限管控、数据脱敏及数据使用审计功能，并贯穿数据存储和应用的全过程。

应用效益

1) 项目经济效益

中联重科工业大数据的应用实践为公司带来了以下经济效益：

降低服务成本：服务成本下降 30%，零配件周转率提升 20%。

提升后市场服务收入：设备租赁服务、二手设备交易以及零配件

销售占比销售额提升 10%。

增值服务收益：深度数据分析带来的增值服务收益提升 30%。

2) 客户预期经济效益

项目实施后，预计将帮助客户提升自身经营管理的能力，为客户降低包括人力、燃油、维修、设备管理等设备运营主要成本。

①通过中联 e 管家为客户提升设备管理效率 30%；

②通过建筑起重机全生命周期管理平台为客户降低安全事故率 20%，设备有效工作时长提升 20%，人力、维修成本降低 30%；

③通过智慧商砼应用提高搅拌站生产效率，新增利润 1.6 亿元/年（按提高 2 万方/站/年，利润 40 元/方，200 家搅拌站测算）。

当前，全球工程机械主要厂商均在尝试工业大数据应用转型，大都处于探索、孵化阶段，尚未形成明显的成功案例和成熟的解决方案。中联重科作为工程机械行业领军企业，其在工业大数据应用方面的积极探索和实践为我国装备制造领域的转型升级起到了示范带头作用，提升了国产工程机械产品的全球竞争力，打造良性生态圈的同时，促进行业向规范化、智能化方向发展。

8、结语

工业大数据价值创造的序幕刚刚开启，不仅在于对现有业务的优化，更在于支撑企业、行业乃至全社会的创新、转型和发展。我们认为，工业大数据应用是价值逐渐加强的过程、是智能制造不断发展的过程、是工业互联网不断普及的过程。

当然，工业大数据又有其鲜明的自身特点。我们必须从这些特点出发，才能找到价值创造的有效途径和方法，不能简单生硬地照搬商业、互联网大数据的做法。应该面向智能制造和工业互联网所带来的创新需求，把工业大数据的应用与工业自身提高质量、降低成本、提升管理水平结合起来，特别要和中国工业的发展阶段和结构特点结合起来，走出有中国特色的工业大数据技术与产业创新路线，助力中国工业弯道取直与加速腾飞。

工业互联网产业联盟
Alliance of Industrial Internet



联系我们

工业互联网产业联盟 秘书处

地址：北京市海淀区花园北路52号，100191

电话：010-62305887

邮箱：aii@caict.ac.cn

网址：<http://www.aii-alliance.org>

