



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

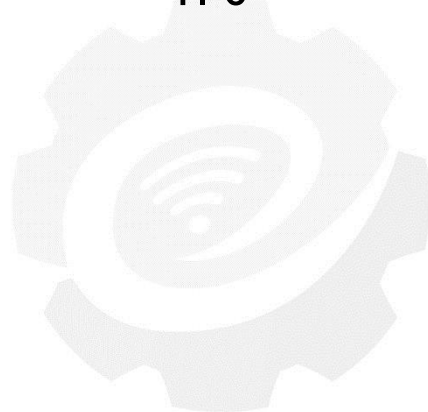
工业大数据技术架构白皮书

(1.0)

工业互联网产业联盟
工业大数据特设组
2018年4月

工业大数据技术架构白皮书

1.0



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

工业互联网产业联盟
工业大数据特设组
2018年5月

声 明

本报告所载的材料和信息，包括但不限于文本、图片、数据、观点、建议，不构成法律建议，也不应替代律师意见。本报告所有材料或内容的知识产权归工业互联网产业联盟所有（注明是引自其他方的内容除外），并受法律保护。如需转载，需联系本联盟并获得授权许可。未经授权许可，任何人不得将报告的全部或部分内容以发布、转载、汇编、转让、出售等方式使用，不得将报告的全部或部分内容通过网络方式传播，不得在任何公开场合使用报告内相关描述及相关数据图表。违反上述声明者，本联盟将追究其相关法律责任。



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

工业互联网产业联盟

联系电话：010-62305887

邮箱：aai@caict.ac.cn

编写说明

党的十九大报告中提出要“加快建设制造强国，加快发展先进制造业，推动互联网、大数据、人工智能和实体经济的深度融合”。再一次强调了运用新兴技术促进信息化和工业化的深度融合，以实现制造强国的战略目标。

工业是国民经济的主导，每一次工业届的重大变革都会对社会发展形成重大的影响。我国政府高度重视并积极推动以互联网为代表的新一代新兴技术与工业系统深度融合，以加速工业体系的智能化变革。工业互联网产业联盟（AII）在2016年发布的《工业互联网体系架构（1.0）》中将工业互联网的建设重点概括为“网络”、“数据”、“安全”三大领域，而“数据”是实现工业智能化的核心驱动。在工业领域中合理地运用大数据技术能有效促进企业信息化发展，提升企业生产运行效率、加速生产信息在制造过程中的流动、助力企业升级转型并形成全新的智能制造模式。

为了加速新一代信息技术与传统产业的融合，工业互联网联盟（AII）针对工业领域的技术创新、标准制定、试验验证、应用实践等进行了一系列调查研究，在工业大数据领域也开展了相关工作，先后发布了《中国工业大数据技术与应用白皮书》，《工业大数据创新竞赛白皮书——风机结冰故障分析指南》等成果，以推动大数据技术在工业领域的深入应用。

本白皮书从实际出发，在现有研究的基础上，结合生产过程中的经典案例，介绍和分析了工业生产环境中大数据技术的应用方法，为工业企业建设大数据系统提供了基础架构层面的建议和指导，从数据的采集与交换、集成与处理、建模与分析、决策与控制几个层面，形成完整的大数据管理与分析架构，供相关行业伙伴参考使用，适用于广义的工业领域，包括制造业、采伐工业、原材料工业以及其他衍生的工业范围。

牵头编写单位：中国信息通信研究院

参与编写单位：清华大学、富士康科技集团、华为技术有限公司、中国联合网络通信集团有限公司、北京理工大学、北京工业大数据创新中心、中国移动通信集团有限公司、工业大数据应用技术国家工程实验室、上海理想信息产业（集团）有限公司、联想（北京）有限公司、天津爱波瑞科技发展有限公司、北京东方国信科技股份有限公司、上海优也商务信息咨询有限公司

编写组成员：

中国信息通信研究院：魏凯、姜春宇、王妙琼、马鹏玮

清华大学：王建民、任良全

富士康科技集团：范治民、徐长恺、李乔如、董吉雄

华为技术有限公司：侯晓钧、殷宏

中国联合网络通信集团有限公司：王非、贺晓伟

北京理工大学：李慧芳

北京工业大数据创新中心：王晨

中国移动通信集团有限公司：毛峻岭、何震宇、刘琨、曹雪峰

工业大数据应用技术国家工程实验室：祝守宇、纪丰伟、郭睿、刘泓汛

上海理想信息产业（集团）有限公司：吴珉杰、余江、尹超迎

联想（北京）有限公司：肖军、黄卓磊、魏星

天津爱波瑞科技发展有限公司：王洪艳、许立红、廖少华、张纪九

北京东方国信科技股份有限公司：赵红卫、梁张伟

上海优也商务信息咨询有限公司：傅源、李志芳、林诗万

法 律 声 明

本报告所载的材料和信息，包括但不限于文本、图片、数据、观点、建议，不构成法律建议，也不应替代律师意见。本报告所有材料或内容的知识产权归工业互联网产业联盟所有（注明是引自其他方的内容除外），并受法律保护。如需转载，需联系本联盟并获得授权许可。未经授权许可，任何人不得将报告的全部或部分内容以发布、转载、汇编、转让、出售等方式使用，不得将报告的全部或部分内容通过网络方式传播，不得在任何公开场合使用报告内相关描述及相关数据图表。违反上述声明者，本联盟将追究其相关法律责任。

工业互联网产业联盟 秘书处

联系电话：010-62305887

联系邮箱：aii@caict.ac.cn

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

目 录

第一章 工业大数据系统综述.....	1
1.1 建设意义及目标.....	1
1.2 重点建设问题.....	2
第二章 工业大数据技术架构概述.....	3
2.1 数据采集与交换.....	5
2.2 数据集成与处理.....	6
2.3 数据建模与分析.....	8
2.4 决策与控制应用.....	9
2.5 技术发展现状.....	10
第三章 工业大数据技术架构实现.....	12
3.1 技术组件选择.....	12
3.1.1 数据采集.....	12
3.1.2 数据存储.....	16
3.1.3 数据计算.....	17
3.1.4 混合云架构.....	18
3.2 建设标准.....	19
3.2.1 基础业务能力.....	19
3.2.2 数据管理能力.....	20
3.2.3 运维管理能力.....	21
3.2.4 安全管理.....	22

3.2.5 性能要求.....	23
3.2.6 开放与兼容性.....	24
第四章 总结与展望.....	25
附录一：典型案例.....	26
一、 富士康 NPI 制造大数据	26
1、 案例背景与业务痛点.....	26
2、 解决方案.....	27
3、 实施效果与推广意义.....	32
4、 案例亮点.....	32
二、 中国联通工业大数据采集分析平台.....	33
1、 案例背景与业务痛点.....	33
2、 解决方案.....	34
3、 实施效果与推广意义.....	37
4、 案例亮点.....	37
三、 华为云 EI 企业智能打造智能化九州通.....	37
1、 案例背景与业务痛点.....	38
2、 解决方案.....	40
3、 实施效果与推广意义.....	45
4、 案例亮点.....	46
四、 潍柴工业大数据平台.....	47
1、 案例背景与业务痛点.....	47
2、 解决方案.....	48
3、 实施效果与推广意义.....	51
4、 案例亮点.....	51

五、 中国移动 OneNET 助力工业资产管理	52
1、 案例背景与业务痛点.....	52
2、 解决方案.....	54
3、 实施效果与推广意义.....	57
4、 案例亮点.....	58
六、 联想工业大数据平台 LEAP.....	59
1、 案例背景与业务痛点.....	59
2、 解决方案.....	59
3、 实施效果与推广意义.....	63
4、 案例亮点.....	64
七、 格力电器工业大数据平台.....	64
1、 案例背景与业务痛点.....	64
2、 解决方案.....	65
3、 实施效果与推广意义.....	69
4、 案例亮点.....	70
八、 合力叉车工业互联网平台.....	71
1、 案例背景与业务痛点.....	71
2、 解决方案.....	76
3、 实施效果与推广意义.....	82
4、 案例亮点.....	82
九、 东方国信大数据助力联合利华能源管理.....	85
1、 案例背景与业务痛点.....	85
2、 解决方案.....	86
3、 实施效果与推广意义.....	90

4、案例亮点.....	90
十、优也基础工业大数据平台在钢铁能效的应用.....	91
1、案例背景与业务痛点.....	91
2、解决方案.....	93
3、实施效果与推广意义.....	95
4、案例亮点.....	97
附录二：参考文献.....	98
附录三：缩略语.....	99



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

第一章 工业大数据系统综述

1.1 建设意义及目标

工业大数据是工业生产过程中全生命周期的数据总和,包括产品研发过程中的设计资料;产品生产过程中的监控与管理数据;产品销售与服务过程的经营和维护数据等。从业务领域来看,可以分为企业信息化数据、工业物联网数据和外部跨界数据。现阶段工业企业大数据存在的问题包括数据来源分散、数据结构多样、数据质量参差不齐、数据价值未有效利用等情况。

工业大数据技术的应用,核心目标是全方位采集各个环节的数据,并将这些数据汇聚起来进行深度分析,利用数据分析结果反过来指导各个环节的控制与管理决策,并通过效果监测的反馈闭环,实现决策控制持续优化。如果将工业互联网的网络比做神经系统,那工业大数据的汇聚与分析就是工业互联网的大脑,是工业互联网的智能中枢。

工业大数据系统的建设首要解决的是如何将多来源的海量异构数据进行统一采集和存储。工业数据来源广泛,生产流程中的每个关键环节都会不断的产生大量数据,例如设计环节中非结构化的设计资料、生产过程中结构化的传感器及监控数据、管理流程中的客户和交易数据、以及外部行业的相关数据等,不仅数据结构不同,采集周期、存储周期及应用场景也不尽相同。这就需要有一个能够适应多种场景的采集系统对各环节的数据进行统一的收集和整理,并设计合理的存储方案来满足各种数据的留存要求。同时需要依据合适的数据治理要求对汇入系统的数据进行标准和质量上的把控,根据数据的类型与特征进行有效管理。之后就需提供计算引擎服务来支撑各类场景的分析建模需求,包括基础的数据脱敏过滤、关联数据的轻度汇总、更深入的分析挖掘等。这些都需要为工业大数据系统选择合适的基础架构作支撑。

建设工业大数据系统能有效地整合工业生产各个环节零散的数据,进行统一的收集、管理和应用,在产品设计环节,全面地收集用户需求,在生产环节有效

地提高设备运行可靠性、提升产品生产效率，在销售环节增强用户体验、提升服务质量，实现制造-服务-用户之间状态和数据的互联互通，使企业能获得全方位的数据感知，及时调整运行模式，做出合理决策。

1.2 重点建设问题

建设企业级工业大数据系统，需要解决多个层面的问题，业务层面需要对各个环节的数据进行梳理和分析，形成完善的数据体系，来描述完整的工业生产流程；技术层面需要建立统一的大数据系统来汇集和处理工业全流程的数据，其中需要根据具体的业务场景选择合适的技术架构，系统建设中需要重点考虑的问题包括以下四个方面：

- 如何采集来自多种数据源的异构数据；
- 如何按照不同的数据留存需求进行高效存储；
- 如何按照业务需求选择数据计算引擎和处理工具；
- 如何保障系统的安全和稳定运行。

本白皮书将围绕这些要点对工业大数据系统的架构做详细的描述。

第二章 工业大数据技术架构概述

本白皮书中描述的工业大数据架构体系，是基于 2016 年工业互联网产业联盟发布的《工业互联网体系架构（1.0）》中对工业互联网数据体系架构描述的细化，见图 1。

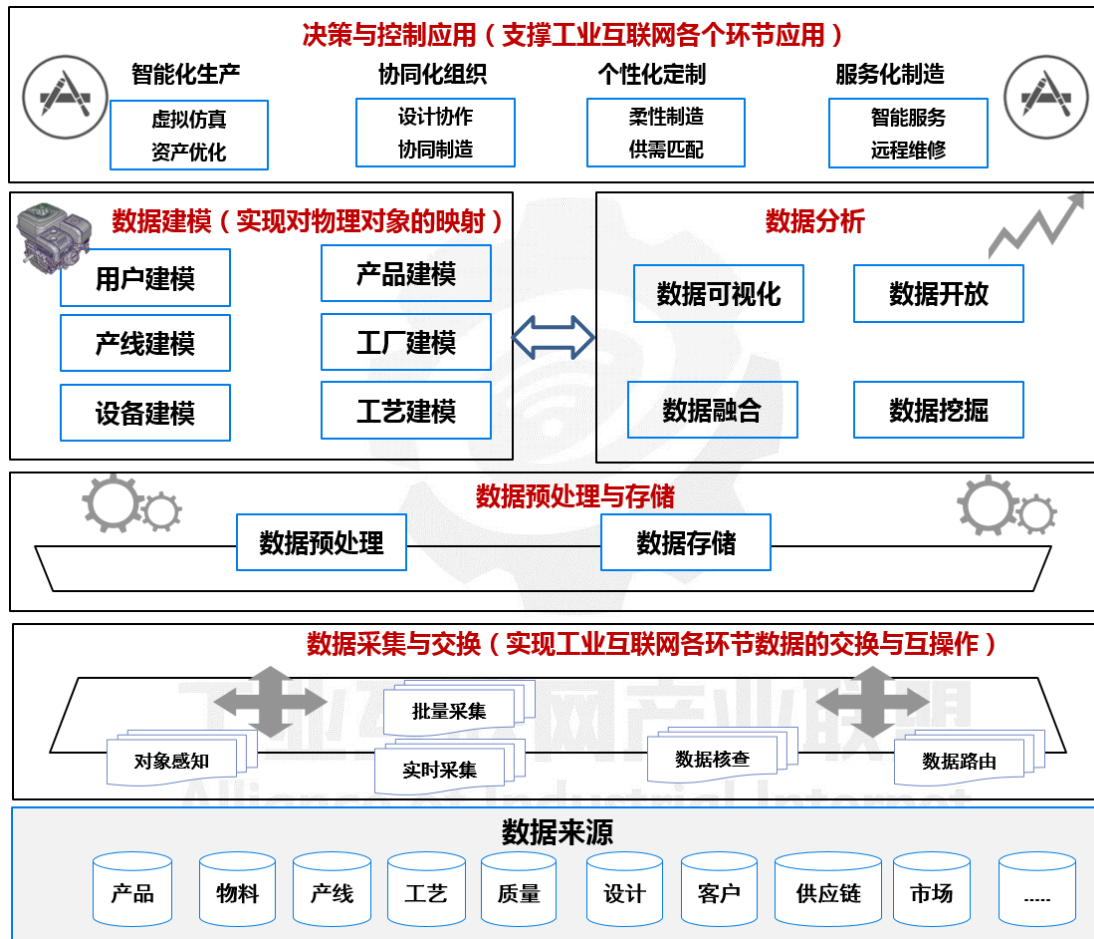


图1 《工业互联网体系架构（1.0）》中的数据体系参考架构

可以总结为数据采集与交换、数据集成与处理、数据建模与分析和数据驱动下的决策与控制应用四个层次，功能架构见图 2。



图2 工业大数据功能架构

采集交换层：主要指从传感器、SCADA、MES、ERP 等内部系统，以及企业外部数据源获取数据的功能，并实现在不同系统之间数据的交互。

集成处理层：从功能上，主要将物理系统实体的抽象和虚拟化，建立产品、产线、供应链等各种主题数据库，将清洗转换后的数据与虚拟制造中的产品、设备、产线等实体相互关联起来。从技术上，实现原始数据的清洗转换和存储管理，提供计算引擎服务，完成海量数据的交互查询、批量计算、流式计算和机器学习等计算任务，并对上层建模工具提供数据访问和计算接口。

建模分析层：功能上主要是在虚拟化的实体之上构建仿真测试、流程分析、运营分析等分析模型，用于在原始数据中提取特定的模式和知识，为各类决策的产生提供支持。从技术上，主要提供数据报表、可视化、知识库、机器学习、统计分析和规则引擎等数据分析工具。

决策控制层：基于数据分析结果，生成描述、诊断、预测、决策、控制等不同应用，形成优化决策建议或产生直接控制指令，从而对工业系统施加影响，实现个性化定制、智能化生产、协同化组织和服务化制造等创新模式，最终构成从数据采集到设备、生产现场及企业运营管理优化的闭环。

功能架构在对应到具体的技术实现时可以参考下图中的技术架构。



图3 工业大数据技术架构

2.1 数据采集与交换

将工业互联网中各组件、各层级的数据汇聚在一起，是大数据应用的前提。要实现数据从底层向上层的汇集，以及在同层不同系统间传递，需要完善的数据采集交换技术支持。工业互联网系统是一个分布式系统，有众多不同的组件组成，为了避免在不同系统间建立连接导致的 N^2 平方复杂性，一般采取消息中间件（Message-oriented middleware）技术来实现。如下图所示，消息中间件的主要功能是实现消息传输管理、队列管理、协议转换等功能。主流消息中间件产品包括 IBM 的 MQ、Oracle 公司的 JMS、微软的 MSMQ 等。消息中间件通过 MQTT、DDS、AMQP、XMPP 等协议与不同系统对接。

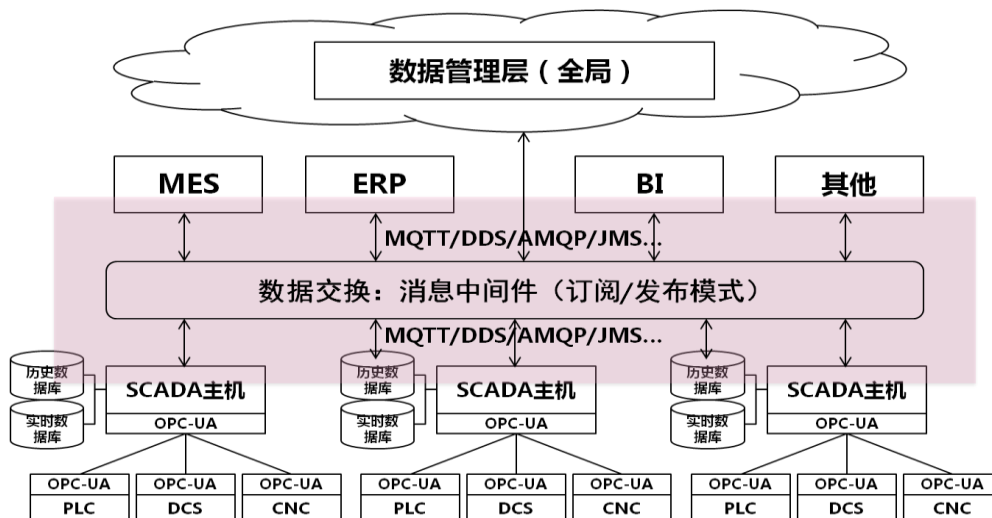


图4 工业大数据采集交换层技术

一个比较大的工业互联网系统中，数据来源和数据需求系统可能有成百上千个，为了简化数据交换，一般采取“发布/订阅”模式传递消息。数据生产者将数据发送给消息中间件，数据消费者则向消息中间件发出一个订阅条件,表示对系统中的哪些数据感兴趣,如果不再感兴趣,则可以取消订阅；而消息中间件则根据一定的路由算法，将生产者发布的事件及时、可靠地传送给所有对之感兴趣的消费者。信息的生产者也称为发布者（Publisher），信息的消费者称为订阅者（Subscriber）。

数据采集是对各种来自不同传感器的信息进行适当转换例如采样、量化、编码、传输。一个数据采集系统，一般包括数据采集器、微机接口电路、数模转换器。

数据交换是指工业大数据应用所需的数据在不同应用系统之间的传输与共享，通过建立数据交换规范，开发通用的数据交换接口，实现数据在不同系统与应用之间的交换与共享，消除数据孤岛，并确保数据交换的一致性。

工业系统中，数据采集与交换是工业系统运作的基底，从微观层每一个零部件信息，到宏观层整个生产流水线信息，如何基于各种网络链接实现数据从微观层到宏观层的流动，形成各个层、全方位数据链条，并保证多源数据在语义层面能够互通，降低数据交换的时延，以实现有效数据交换，技术上是一个比较大的挑战。

2.2 数据集成与处理

工业大数据集成就是将工业产品全生命周期形成的许多个分散的工业数据源中的数据，逻辑地或物理地集成到统一的工业数据集合中。工业大数据集成的核心是要将互相关联的分布式异构工业数据源集成到一起，使用户能够以透明的方式访问这些工业数据源，达到保持工业数据源整体上的数据一致性、提高信息共享与利用效率的目的。

工业大数据处理是利用数据库技术、数据清洗转换加载等多种工业大数据处理技术，将集成的工业数据集合中大量的、杂乱无章的、难以理解的数据进行分

析和加工，形成有价值、有意义的数据。

工业大数据集成处理层，主要涉及数据的抽取转换加载（ETL）技术、数据存储管理技术、数据查询与计算技术，以及相应的数据安全管理和数据质量管理等支撑技术。其中，ETL、数据查询与计算等技术，与互联网大数据技术相似，而基于开源的 Hadoop 等技术将成为未来的发展趋势，具体如图 4 所示。

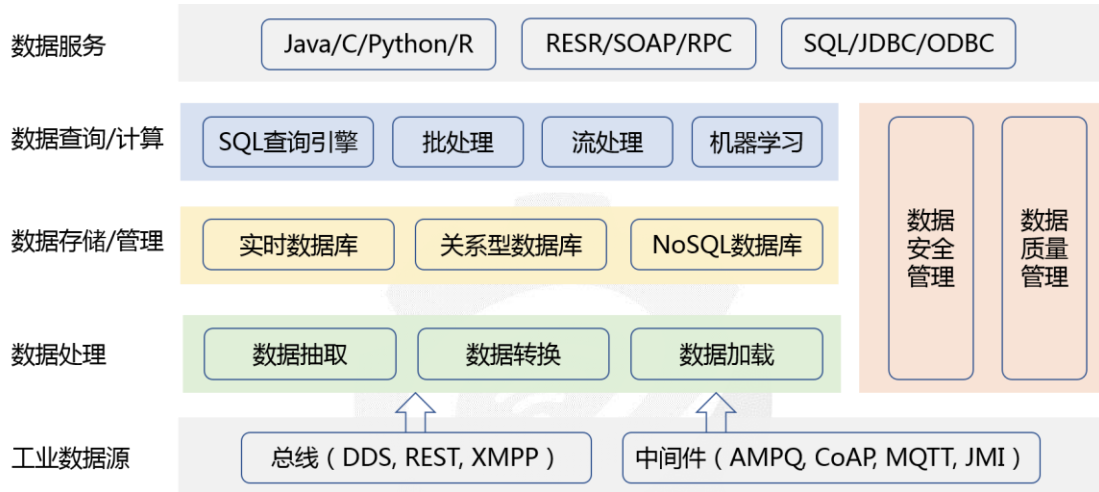


图5 工业数据集成处理框架

其中，ETL 包括三部分：数据抽取、清洗转换与加载。数据抽取主要将分散的、异构工业数据源中的数据如关系数据、平面数据文件等抽取到临时中间层；数据清洗是对抽取到临时中间层的数据进行审查、过滤和校验，旨在去除噪声数据、删除重复信息、纠正错误，并维护数据的一致性；数据转换主要包括数据格式规范化与数据拆分等，数据规范化实现字段格式的约束定义，以利于数据的建模与分析；数据加载是将已经加工好的数据加载到数据仓库中。

由于加载到数据仓库中数据量巨大，且包含结构化、半结构化和非结构化数据，传统的关系型 SQL 数据库难以满足大数据的存储与管理。因此，需要借助实时数据库、关系数据库、NoSQL 数据库，实现工业大数据的存储与管理。实时数据库是基于实时数据模型建立，用于处理不断更新、快速变化以及具有时间限制的数据，随着技术的演进，时序数据库也逐渐兴起，在部署方式、检索性能及使用成本上对比传统实时数据库均有优化；关系型数据库是采用关系模型来组织数据，用于处理永久、稳定的数据；NoSQL 数据库是指非关系型的数据库，

具有灵活的可扩展性，在大数据量下具有非常高的读写性能。

数据查询与计算主要采用 SQL 查询引擎、批处理、流处理、机器学习等方法。其中，SQL 查询引擎将用户输入的 SQL 语句序列转换为一个可执行的操作序列，并返回查询结果集；批处理主要操作大容量静态数据集，并返回计算结果；流处理则对实时进入系统的数据进行计算，处理结果即时可用，并将随新数据的到达持续更新。

数据服务层的主要作用是提供数据服务的接口，以实现工业大数据的访问、更新等基本功能。

2.3 数据建模与分析

数据建模是根据工业实际元素与业务流程，在设备物联数据、生产经营过程数据、外部互联网等相关数据的基础上，构建供应商、用户、设备、产品、产线、工厂、工艺等数字模型，并结合数据分析提供诸如数据报表、可视化、知识库、数据分析工具及数据开放功能，为各类决策提供支持。工业大数据分析建模技术，已经形成了一些比较成熟稳定的模型算法。从大的方面可以分为基于知识驱动的方法和基于数据驱动的方法。有时候数据可视化技术本身也被称为一种数据分析方法。工业大数据建模分析技术体系如图 5 所示。

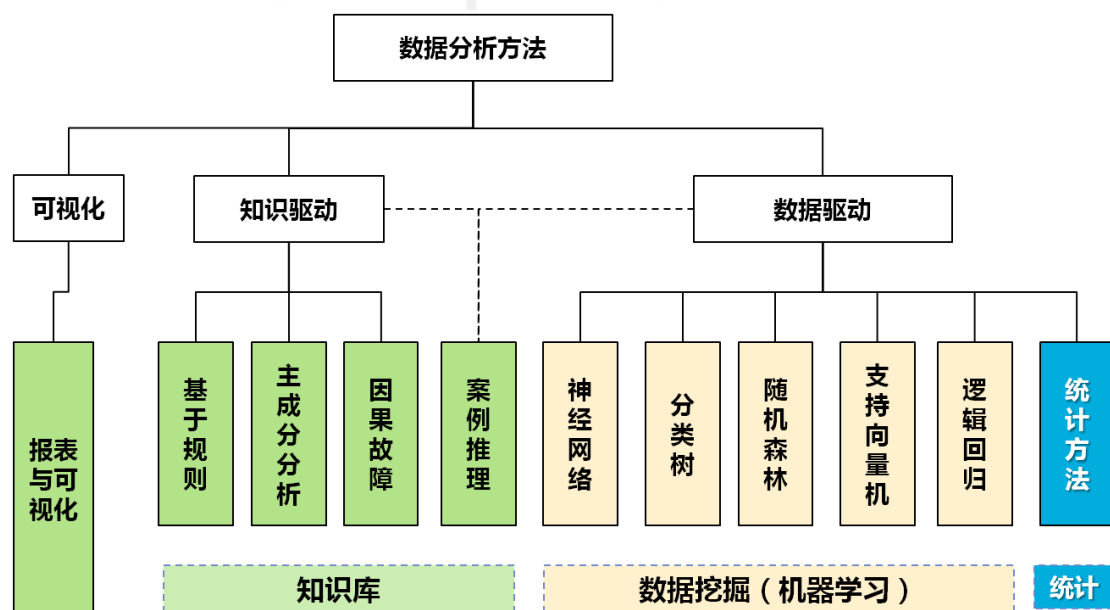


图6 工业大数据分析技术体系

知识驱动的分析方法，是基于大量理论模型以及对现实工业系统的物理、化学、生化等动态过程进行改造的经验，建立在工业系统的物理化学原理、工艺及管理经验等知识之上，包括基于规则的方法、主成分分析技术、因果故障分析技术和案例推理技术等。其中，知识库是支撑这类方法的基础。

数据驱动的分析方法，很少考虑机理模型和闭环控制逻辑的存在，而是利用算法在完全数据空间中寻找规律和知识，包括神经网络、分类树、随机森林、支持向量机、逻辑回归、聚类等机器学习方法，以及基于统计学的方法。两类方法的对比如表 1 所示，分别适用于不同的场景。

表 1：工业数据分析的两类主要方法对比

方法	知识驱动的分析方法	数据驱动的分析方法
优势	1.有理论基础和确定的因果关系 2.直观，容易理解 3.数据量和计算量需求小	1.可以动态调整 2.对先验知识的依赖程度低 3.建模周期短
劣势	1.是静态模型，离线无法动态调整 2.创新理论和模型建立周期比较长	1.数据量和计算量需求大 2.因果关系不明确，可解释性较差 3.置信度难以达到工业级要求

2.4 决策与控制应用

根据数据分析的结果产生决策，从而指导工业系统采取行动，是工业大数据应用的最终目的。工业大数据应用可以分为以下 5 大类：

(1) 描述类 (descriptive) 应用：主要利用报表、可视化等技术，汇总展现工业互联网各个子系统的状态，使得操作管理人员可以在一个仪表盘 (dashboard) 上总览全局状态。此类应用一般不给出明确的决策建议，完全依靠人来做出决策。

(2) 诊断类 (diagnostic) 应用：通过采集工业生产过程相关的设备物理参数、工作状态数据、性能数据及其环境数据等，评估工业系统生产设备等运行状态并预测其未来健康状况，主要利用规则引擎、归因分析等，对工业系统中的故障给出告警并提示故障可能的原因，辅助人工决策。

(3) 预测类 (predictive) 应用：通过对系统历史数据的分析挖掘，预测系统的未来行为。主要是利用逻辑回归、决策树等，预测未来系统状态，并给出建议。

(4) 决策类 (deceive) 应用：通过对影响决策的数据进行分析与挖掘，发现决策相关的结构与规律，主要是利用随机森林、决策树等方法，提出生产调度、经营管理与优化方面的决策建议。

(5) 控制类 (control) 应用：根据高度确定的规则，直接通过数据分析产生行动指令，控制生产系统采取行动。

基于大数据的工业决策控制技术的框架如下图所示。

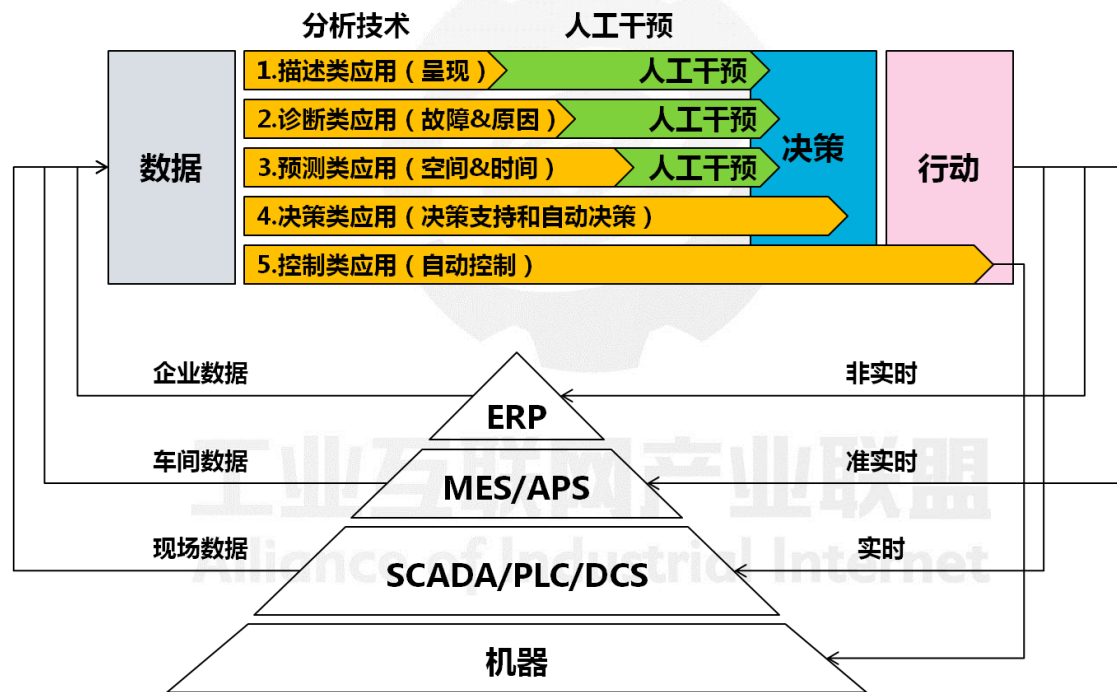


图7 工业大数据决策与控制应用技术

2.5 技术发展现状

现阶段工业大数据的建设仍处于发展极不均衡的状态，部分先进的工业企业已经利用大数据技术整合各环节数据资源，为生产和服务提供精准的数据支撑，但仍有部分工业企业尚未实现信息化或信息化程度不高，需要从最底层的设备开始更新与改造，以适应新的数字化生产流程，规范各环节数据资料，结合大数据技术进一步地优化生产过程和服务质量。在发展工业大数据的各阶段下主要出现

了以下三个问题：

(1) 设备老旧导致数据采集困难：在许多制造行业中，由于仍然使用传统的生产设备，并且因为各类原因对产线设备进行更新换代较为困难，一些关键的生产数据无法采集，不能全面的监控产线的生产情况。这种情况下，需要在设备上加装外置传感器，并对传感器采集到的生产数据进行实时地汇集监控，来全面感知产线的生产情况。

(2) 数据跨平台跨部门整合耗时：在工业领域进行业务分析时，往往需要整合多个数据源的信息，这些数据种类繁多、数据量大、数据质量参差不齐，有时甚至需要跨部门沟通，走繁琐的审批流程，在经过了一系列关卡收集并整理好数据后，由于耗费时间太长，一些具有时效性的数据就会产生价值流失。需要建设持续的、自动化的、全量的数据采集机制，来保证数据的价值被有效地利用。

(3) 初期缺乏规划导致应用竖井：在开展大数据应用的初期，由于大部分是业务驱动，各项目单独建设大数据系统，缺乏统一规划，容易造成应用竖井，数据重复采集、基础信息不一致，从而不可避免的出现资源浪费、数据关联性差，无法高效的利用数据价值。所以需要企业有全面的大数据系统建设规划，对数据进行统一采集和管理，为后续各种场景的分析应用提供良好的数据环境。

第三章 工业大数据技术架构实现

3.1 技术组件选择

3.1.1 数据采集

随着工业制造中各层的精细化与制程的高密度化，工业制造所产生的数据必然形成好几个等级的增长。面对如此庞大与多样的数据整合问题，企业必须有整体统一的数据汇聚与应用策略，设计通用可靠的数据采集机制，来满足各方面的数据采集需求。数据采集的完整性、准确性，决定了数据应用是否能真实可靠地发挥作用。因此，在建设数据采集系统时，建议着重考虑以下五个要求：

(1) 数据接口通用性：由于新技术更新换代较频繁，需要进行版本管理，并定期更新接口，建议用统一的数据交换格式来适应接口频繁更新的情况，使接口能够快速调整。

(2) 支持广泛的数据源：采集技术需支持尽可能多的数据源端。

(3) 支持横向扩展：当设备增加时，所造成的性能瓶颈须能通过横向扩展的方式解决。

(4) 保证数据不遗失：采集过程中须确保数据准确地、不遗失地送达处理层与储存层。

(5) 避免增加基础建设复杂度：在不断扩充设备的情况下，采集技术不应增加基础建设扩展时的负担。

从工业数据的来源进行分类，主要包括管理系统、生产系统、外部数据三大方面的数据来源。从数据采集的全面性上看，不仅要涵盖基础的结构化交易数据，还将逐步包括半结构化的用户行为数据，网状的社交关系数据，文本或音视频类型的用户意见和反馈数据，设备和传感器采集的周期性数据，以及未来越来越多有潜在意义的各类数据。下表整理出了一些工业大数据系统中常见的数据源及其数据特性，供参考：

表 1：常见工业数据源分类

分类	系统类型	典型系统	数据结构	数据特点	实时性
管理系统	设计资料	产品模型、图纸文档	半结构化/非结构化	类型各异、更新不频繁、是企业核心数据	批量导入
	价值链管理	供应链 SCM、客户关系 CRM	结构化/半结构化	没有严格的时效性要求，需要定期同步	批量导入
	资源管理	ERP/OA、MES、PLM、环境管理系统、仓库管理系统、能源管理系统	结构化	没有严格的时效性要求，需要定期同步	批量导入
生产系统	工业控制系统	DCS、PLC	结构化	需要实时监控，实时反馈控制	实时采集
	生产监控数据	SCADA	结构化	包含实时数据和历史数据	实时采集/批量导入
	各类传感器	外挂式传感器、条码、射频识别	结构化	单条数据量小，并发度大，结合 IoT 网关	实时采集
	其他外部装置	视频摄像头	非结构化	数据量大、低时延，要求网络带宽和时延	实时采集
外部数据	外部数据	相关行业、法规、市场、竞品、环境数据	非结构化	数据相对静止，变化较小，定期更新	批量导入

管理系统数据采集：这里讨论的管理系统的数据包括了工业产品的设计资料、价值链管理数据及生产过程中的资源管理数据。

1.设计资料：设计资料大多来源于传统工业设计和制造类软件，如：CAD、CAM、CAE、CAPP、PDM 等。这类数据主要是各类产品模型，以及相关的图纸或电子文档，大多数为非结构化数据。这些设计类数据的采集对时效性要求不高，只需定期批量导入大数据系统。

2.价值链管理数据：价值链数据主要指企业生产活动中上下游的信息流数据，主要来源于供应链管理系统（SCM）、客户关系管理系统（CRM）等。这类数据主要包含供应链信息和客户信息，通常是规范的结构化数据，采集时对时效性要求不高，只需按业务分析要求的更新周期定期批量导入大数据系统。

3.资源管理数据：资源管理数据的来源主要是生产环节的各类管理系统，包括企业资源计划（OA/ERP）、生产过程执行系统（MES）、产品生命周期管理

(PLM)、环境管理系统 (EMS)、仓库管理系统 (WMS)、能源管理系统等。这类数据主要描述了生产过程中的订单数据、排程数据、生产数据等，大多数为标准的结构化数据，采集时对时效性要求不高，只需按业务分析要求的更新周期定期批量导入大数据系统。

生产系统数据采集：这里讨论的生产系统数据主要来自工业控制系统、生产监控系统、各类传感器以及其他外部装置。

1.工业控制系统数据：工业控制系统数据的来源主要包括分布式控制系统 (DCS)，以及可编程逻辑控制器 (PLC) 这类系统。通常 DCS 与 PLC 共同组成本地化的控制系统，主要关注控制消息管理、设备诊断、数据传递方式、工厂结构，以及设备逻辑控制和报警管理等数据的收集。此类数据通常为结构化数据，且数据的应用通常对时效性要求较高，需要数据能及时地上报到上层的处理系统中。

2.生产监控数据：生产监控数据主要来源于以 SCADA 为代表的监视控制系统。SCADA 系统的设计用来收集现场信息，将这些信息传输到计算机系统，并且用图像或文本的形式显示这些信息。这类数据也是规范的结构化数据，但相对 DCS 和 PLC 系统来说，SCADA 系统可以提供实时的数据，同时也能提供历史数据。因此在考虑数据的采集策略时，需要根据上报数据的类型来选择是实时采集或是批量导入。

3.各类传感器：在生产车间的很多生产设备并不能提供生产数据的采集和上传，因此需要通过外接一套额外的传感器来完成生产数据的采集。外挂式传感器主要用在无生产数据采集的设备或者数据采集不全面的设备上、以及工厂环境数据的采集。同时外挂式传感器根据使用现场的需求，可以采用接触式的传感设备和非接触式的传感设备。此类数据的单条数据量通常都非常小，但是通信总接入数非常高，即数据传输并发度高，同时对传输的实时性要求较高。

4.其他外部装置：其他外部装置产生的数据以视频摄像头为例，数据主要来源于对产品的质量监控照片、视频，或者是工厂内的监控视频等。此类数据的特点是数据量大，传输的持续时间长，需要有高带宽、低时延的通信网络才能满足

数据的上传需求。对于其他不同于视频数据的外部装置数据，需要针对数据的特性进行采集机制的选择。

外部系统数据采集：外部系统数据主要来源于评价企业环境绩效的环境法规、预测产品市场的宏观社会经济数据等，此类数据主要用于评估产品的后续生产趋势、产品改进等方面，与管理系统的信息采集类似，可以通过标准的 RJ45 接口进行数据的传输。通常本类数据相对静止，变化较小，因此数据的上传频次较低。

综合上述多类数据源的采集场景和要求，系统的集成导入应同时具备**实时接入**（如：工业控制系统、生产监控数据、各类传感器）和**批量导入**（如管理系统、外部数据）的能力，同时能根据需要提供可定制化的 IoT 接入平台。具体建设要求如下：

（1）对于需要实时监控、实时反向控制类数据，可通过实时消息管道发送，支持实时接入；如：工业控制系统数据、生产监控系统数据等。建议可采用如 Kafka、Fluentd 或是 Flume 等技术，这类技术使用分布式架构，具备数据至少传输一次的机制，并为不同生成频率的数据提供缓冲层，避免重要数据的丢失。

（2）对于非实时处理的数据，可采取定时批量地从外部系统离线导入，必须要支持海量多源异构数据的导入；如：资源管理数据、价值链数据、设计资料等。建议可采用 Sqoop 等数据交换技术，实现 Hadoop 与传统数据库（MySQL、Oracle、Postgres 等）间大批量数据的双向传递。

（3）当系统中有大量设备需要并发且多协议接入时，如各类传感器件，可部署专业 IoT 接入网关，IoT 接入平台需同时具备支持 TCP、UDP、MQTT、CoAP、LWM2M 等多种通信协议的能力。在面对各类传感器的数据采集时，可以结合 RFID、条码扫描器、生产和监测设备、PDA、人机交互、智能终端等手段采集制造领域多源、异构数据信息，并通过互联网或现场总线等技术实现源数据的实时准确传输。有线接入主要以 PLC、以太网为主。无线接入技术种类繁多，包括条形码、PDA、RFID、Zigbee、WiFi、蓝牙、Z-wave 等短距离通信技术和长距无线通信技术。其中，长距离无线技术又分为两类，包括工作于未授权频谱的 LoRa、SigFox 等技术和工作于授权频谱下传统的 2/3/4G 蜂窝技术及其 3GPP 支持的 LTE

演进技术，如 LTE-eMTC、NB-IOT 等。

3.1.2 数据存储

工业大数据系统接入的数据源数量大类型多，需要能支持 TB 到 PB 级多种类型数据的存储，包括关系表、网页、文本、JSON、XML、图像等数据库，应具备尽可能多样化的存储方式来适应各类存储分析场景，总结为如下表格：

表 2：各类存储对应适用场景

类型	典型介质	适用场景
海量低成本存储	对象存储、云盘	海量历史数据的归档和备份
分布式文件系统	HDFS、Hive	海量数据的离线分析
数据仓库	MPP、Cassandra	报表综合分析、多维随机分析
NoSQL 数据库	HBase、MongoDB	各类报表文档，适用于简单对点查询及交互式查询场景
关系型数据库	MySQL、SQLServer、Oracle、PostgreSQL	适用于交互式查询分析
时序数据库	InfluxDB、Kdb+、RRDtool	依据时间顺序分析历史趋势、周期规律、异常性等场景
内存数据库	Redis、Memcached、Ignite	数据量不大且要求快速实时查询场景
图数据库	Neo4j	分析关联关系及具有明显点/边分析的场景
文本数据索引	Solr、Elasticsearch	文本/全文检索

在不同的工业数据应用场景中，数据存储的介质选择十分重要，下面列举一些经典的使用场景来介绍如何选择存储技术：

(1) 实时监控数据展示：通常情况下实时采集的监控数据在进行轻度的清洗和汇总后会结合 Web UI 技术实时展现生产线的最新动态。这类及时性互动性高的数据一般使用内存数据进行存储，如 Redis、Ignite 等技术，可以快速响应实时的查询需求。

(2)产线异常的分析与预测：使用机器学习技术对产线数据进行深入挖掘分析运行规律，可以有效地对产线的异常进行分析和预测，进而改善制程、减少损失、降低成本及人为误判的可能性。这类用于分析的历史数据一般选择使用HDFS、Cassandra等分布式储存，适用于海量数据的探索和挖掘分析。同时，对于这类与时间顺序强相关的分析场景，数据的存储可以选择 InfluxDB 这类时序数据库，可以极大提高时间相关数据的处理能力，在一定程度上节省存储空间并极大地提高查询效率。

(3)商业智能：如果需要整合多种数据来制作商业策略性报表，适合使用结构化储存，比如传统的关系型数据库，MySQL、Oracle等。如果需要考虑性能和及时性，可以考虑分类存储至 NoSQL 数据库，如 Cassandra、HBase 与 Redis 等。

3.1.3 数据计算

大数据系统通常需要能够支持多种任务，包括处理结构化表的 SQL 引擎、计算关系的图处理引擎和进行数据挖掘的机器学习引擎，其中面向 SQL 的分析主要有交互式查询、报表、复杂查询、多维分析等。

表 3：各类计算引擎对应适用场景

类型	典型介质	适用场景
实时计算引擎	Storm、Spark Streaming、Flink	设备监控、实时诊断等对时效性要求较高的场景
离线计算引擎	MapReduce、Spark、Hive	适用于大数据量的，周期性的数据分析，例如阶段性的营销分析，或生产能耗分析等
图计算引擎	Graphlab、GraphX	适用于事件及人之间的关联分析，比如建立用户画像进行个性化定制或营销
数据综合分析 OLAP	MPP	产线或销售环节的综合报表分析
业务交互查询 OLTP	MySQL、SQLServer、Oracle	交互式查询分析
分布式数据库中间件	Cobar、TTDL、MyCAT	海量数据高并发时的弹性扩容解决方案
数据挖掘能力	Spark、TensorFlow	需要迭代优化的数据挖掘场景，如故障预测、用户需求挖掘等

1.实时计算引擎，包括 Storm、Spark Streaming、Flink 等业界通用架构，适用于基于窗口或消息的实时数据处理，结果响应的时延要求在毫秒级；

2.离线计算引擎，包括 MapReduce、Spark、Hive，适用于批数据分析和定时分析等；

3.图计算引擎，适用于事件及人之间的关联关系分析；

4.数据综合分析 OLAP，如 MPP 数据库，适用于综合报表分析；

5.业务交互查询 OLTP，如 MySQL、SQLServer、Oracle、PostgreSQL 等，适用于交互式查询分析；

6.分布式数据库中间件，可解决数据库容量、性能瓶颈和分布式扩展问题，提供分库分表、读写分离、弹性扩容等能力，适用于海量数据的高并发访问场景，有效提升数据库读写性能。

7.数据挖掘能力，为了能够匹配工业大数据决策与控制应用的 5 大场景，特别是诊断类、预测类、决策类应用闭环的要求，系统应该具备完善的机器学习、深度学习、图计算等平台级能力。机器学习能力如基于开源 Spark 框架推出的算法库 MLlib、GraphX 等；深度学习有 TensorFlow、Caffe、MXNet 等平台；图计算能力，业界相对比较流行的开源产品有 Titan，另外还有很多优秀的商业产品可供选择。

总体来说，大数据平台的计算组件应该能够支持批量和实时两大类任务，同时具备精细化的任务和资源调度的能力。

3.1.4 混合云架构

结合工业企业的 IT 现状和对数据安全、建设成本等因素的综合考量，可以引入混合云架构来满足现代工业大数据建设的诉求。对实时性要求高，与生产强相关，特别是需要及时闭环控制的应用系统可部署在线下，而大数据量的分析类、预测类应用可以部署在云上，尤其是偏物联网的应用，这样可以有效均衡架构的私密性、便捷性、可维护性及性价比。

表 4：部署选型建议

类型	建议
本地化部署	控制系统、ERP、MES 等需要实时反馈或者对数据安全要求较高的适合本地化部署
云上部署	偏物联网相关分析类、预测类应用可以选择云上部署

3.2 建设标准

3.2.1 基础业务能力

首先要考虑工业大数据系统功能的完整性，即支撑大数据应用全生命周期的基础业务能力，例如接入、存储、分析等。基础业务能力的考虑方向主要包括数据导入、存储与计算、多任务引擎等三方面。

- 数据导入。大数据系统必须要支持海量多源异构数据导入，具体来说需要支持传统数据库、本地、FTP 等多种数据源；支持结构化、半结构化和非结构化数据的导入；支持定时、实时、循环任务的数据导入方式。
- 数据标准化。系统需要提供能够对数据进行有效处理和管理的工具能力，使进入系统的数据符合企业的数治理要求，保证平台数据的完整性、有效性、一致性、规范性、开放性和共享性。
- 数据存储和计算。大数据平台应该能支持 TB 到 PB 级多种类型数据的存储，包括关系表、网页、文本、JSON、XML、图像等数据库。平台的计算组件应该能够支持批量和实时两类任务，同时具备精细化的任务和资源调度的能力。
- 多任务引擎。数据平台需要能够支持多种任务，包括处理结构化表的 SQL 引擎、计算关系的图处理引擎和进行数据挖掘的机器学习引擎。其中面向 SQL 的分析主要有交互式查询、报表、复杂查询、多维分析等。
- 基础分析模型。大数据系统应具备基础的业务分析模型，能够针对特定场景的分析要求，进行自动化的业务自助分析。

- 可视化报表工具。大数据系统应能提供生产可视化报表的能力，需要提供常用的折线图、柱状图、饼图、表格等组件，并支持自定义可视化组件或第三方可视化工具。

3.2.2 数据管理能力

工业大数据系统的基础数据管理能力¹应包括以下几项：

- 数据标准制定：工业大数据系统需要支持统一的数据标准制定，使用合理的数据标准，可以有效约束平台数据的完整性、有效性、一致性、规范性、开放性和共享性，从而提高企业进行数据治理的水平。
- 数据模型管理：数据模型是对数据特征的抽象，用于描述一组数据的概念和定义。大数据系统中的数据模型管理应支持数据模型的设计、数据模型和数据标准词典的同步、数据模型的审核发布、差异对比、版本管理等。能有效指导企业进行数据整合，提高数据质量。
- 元数据管理：元数据是描述数据的数据。大数据系统中的元数据管理能对数据进行有效地解释说明并有助于企业理解数据的真实含义。
- 数据质量管理：数据质量是保证数据应用的基础。大数据系统中的数据质量管理机制需要能保证数据的完整性、规范性、一致性、准确性、唯一性和关联性，来帮助企业获得高质量的、结构清晰的数据，以更好的服务上层应用。
- 生命周期管理：生命周期管理是指对数据产生、存储、传输、使用和删除的全过程进行管理，依据不同数据在不同阶段的价值实施不同的管理策略，降低存储成本，提升数据价值，以达到最高效的管理效果。

¹了解更多行业通用数据治理要求可参考数据中心联盟发布的《数据管理实践白皮书（1.0版）》

- **数据安全**管理：大数据系统应具备针对数据的安全管理策略，从隐私保护、信息加密、鉴权控制、日志审计等多个方面确保数据安全，做到事前可管、事中可控、事后可查。
- **数据开放**：数据开放主要指基于数据资源，开展数据共享和交换，通过各种管控机制的保障，使数据能通过标准化接口方式提供给外部需求方，发挥更大的价值。

3.2.3 运维管理能力

大数据平台在生产环境下的部署、运行与维护，需要做到高可靠、简操作、易扩展，避免后期维护产生高昂成本。需要从大数据平台的运维能力、弹性扩展能力和安全防护能力等几个维度考虑。

- **运维能力**。支持一键式或者向导式的安装部署；支持集群平滑升级；能够对集群、各类组件、任务状态进行监控，进行启动、停止、增加、卸载等常规操作，并能够配置集群的各项参数；能够收集集群和组件的运行的日志，对日志进行检索和下载；能通过界面、邮件、短信等形式对集群的各类故障进行告警，能够在界面对告警值域进行配置；支持运维用户的角色分类，支持用户账号的增、删、改，细粒度的权限分配。
- **弹性扩展能力**。大数据系统需要能随着数据和业务的快速发展而自由扩展，可扩展性是大数据平台的重要能力之一，一是要支持集群的水平与垂直扩展，提升大数据平台的存储和计算能力，二是实现数据的快速分布和自动均衡，无需人工过多干预。
- **高可用**。大数据系统需要支持包括数据节点、服务节点、网络环境的主备切换能力，从而保证服务的延续性。
- **备份管理**。大数据系统需要具备风险预防机制和灾难恢复措施，系统中的数据需要按照不同类别进行不同周期、不同方式和地理位置的区分备份。

3.2.4 安全管理

安全管理的目的是为了保证系统安全运行，与此同时防止系统受到外来攻击、破坏和非法访问，需要在不同层次利用多种手段来保证系统的安全。安全管理主要包括系统的主机安全、网络安全、数据安全、应用安全、以及数据访问审计日志等功能。

安全指标考量大数据系统是否能够提供基本的安全方案，以防止恶意的访问和攻击，防止关键数据的泄露，可以从以下几个方面考量：

- 主机安全：大数据系统需要选择安全的操作系统版本，并对操作系统进行基础的安全配置和安全加固，以确保系统安全、可靠、高效地运行。
- 网络安全：网络系统和服务器系统具有入侵检测的功能，可监控可疑的连接、非法访问等，采取的措施包括实时报警、自动阻断通信连接或执行用户自定义的安全策略。网络和服务器系统能定期检查安全漏洞及病毒，根据扫描的结果更正网络安全漏洞和系统中的错误配置；使用加密技术对在互联网上传输的重要数据进行加密。与外部系统连接配置防火墙设备，并定义完备的安全策略。
- 数据安全：数据安全的保证数据库和其它文件只能被授权用户访问和修改，防止在本地存储或者网络传输的数据受到非法篡改、删除和破坏。数据相关的安全控制包括数据加密、访问控制、数据完整性、数据防篡改。
- 应用安全：需要对账号进行集中管理和统一认证，并对操作进行记录和审计，防范 SQL 注入、防范跨站攻击等。
- 日志审计：对设备日志、操作系统日志、系统平台日志、应用日志等进行留存和审计。

3.2.5 性能要求

需要全面考察平台在不同数据规模和任务场景下的性能表现。主要的指标有吞吐量、响应时间、最大并发等。实际性能表现需要对平台进行测试，典型的测试场景包括根据自身业务确定的单项任务和多种混合任务测试，以及压力测试和稳定性测试。考虑到大数据系统运营的基本需求，需要考察以下性能项：

平台性能	计算引擎处理能力
	请求响应时延
	支持用户数
	支持并发数
数据抽取性能	准实时数据更新时延
	日增量数据更新时延
	月数据更新时延
网络传输性能	网络传输速度
数据导出性能	数据导出吞吐量
可靠性性能	主备机保证系统 7*24 小时不间断工作
	告警通知时延
	每年例外停机时间
	主备机切换间隔
	系统平均无故障时间

3.2.6 开放与兼容性

大数据系统的建设，还需要考虑到开放性与兼容性，能够与既有系统无缝衔接，能够兼容支持各类数据源、外围协同系统及上层各类应用。

- 开放性。要求能够支持主流的开源技术，比如 Hadoop、Spark、MySQL、Greenplum 等开源社区技术，能够对相关的组件进行替换和更新，方便集成与优化。同时提供开放接口，支持与各类外部系统的对接。
- 兼容性。由于传统用户的大部分数据分析任务是以结构化数据为主的 SQL 任务，为了节约学习成本，实现平稳过渡，大数据系统要求能兼容更多的 SQL 的标准和语法；其次需要支持 JDBC、ODBC 等通用接口，从而保证对接传统的数据库、上层 BI 工具等各类上下游产品，方便系统和应用开发的便捷性；系统还需要能够支持异构的硬件和不同的操作系统，从而保证上层应用对于异构软硬件设备透明能力，充分利用各类资源。

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

第四章 总结与展望

随着对工业大数据体系架构研究的不断深入，我们将现阶段工业大数据系统建设的几个要点列举如下：

一是跨层次跨环节的数据整合。当前的工业数据以水平来看分散在研发设计、生产管理、企业经营等各个环节，垂直来看分散在生产现场、企业管理等不同层次，为了更好地支撑生产经营各环节的应用分析需求，需要更深入完整地进行数据整合。

二是选择合适的架构进行系统的部署及数据的集成。大数据系统的实施需要充分考虑企业现状，以及对数据安全、建设成本等多项因素的要求，有效结合本地化服务与云服务，以达到最佳的使用效果。

三是建立适合企业的资产管理机制，将数据看作企业资产，梳理并明确各类数据的意义，形成企业级甚至行业级的通用数据模型，制定数据安全规范，对数据进行高效管理，提升数据质量，以实现更大的价值。

四是关联工业生产各环节的数据进行分析。工业生产流程和经营流程环环相扣，每一步的变动都可能影响到下一步的发展，数据分析也应与工业系统的各环节紧密结合，以形成闭环的设备、生产现场及企业运营管理优化的整体态势分析。

本白皮书描述了工业大数据系统在建设初期最基础的技术架构要求，而工业企业建设大数据系统的最终目的是为了解决生产和经营中的实际问题，为企业切实地节省资源成本，带来利润地增长。大数据价值的实现是很长的一段路，需要我们不断地进行摸索，使数据真正地运用起来。在互联网时代，各家企业不仅需要在技术上深耕，更要以开放的眼光看待整个产业，积极地贡献和分享建设经验，使好的案例能在行业快速的复制推广，在实现自身价值的同时，加速推进整个产业的变革。

附录一：典型案例

一、 富士康 NPI 制造大数据

1、 案例背景与业务痛点

新产品制造在 NPI (New Product Introduce, 新产品导入) 阶段需在有限的时间内使得每个步骤都达到客户的规格要求。整个 NPI 的生命周期可以分成：

- (1) Proto 原型机阶段
- (2) EVT 工程验证测试阶段
- (3) DVT 设计验证测试阶段
- (4) PVT 生产验证测试阶段
- (5) Ramp 快速爬坡
- (6) MP 量产阶段

也因此，在每一个阶段工程单位与测试单位都需要完整的生产信息与关键物料信息的整合，协助 FA (Failure Analysis) 工程师与 RD 工程师合作，进行问题分析与解决方案设计。

由于每一个阶段都是分秒必争，并且还要能快速反应问题，求新求变，必须在限定时间内分析所有的问题且找出原因（当天问题当天解决）这将是一个高度困难的挑战与目标。而为能加快 FA 工程师分析的速度以及有更多的时间做进阶数据分析，快速收集所有生产信息，甚至往前追溯上游供应链关键信息都是非常必要的，但却又相当耗费时间。

依据过去的经验，工程师通常会花费 80% 的时间做数据收集与整合，却只用 20% 的时间做数据分析与问题诊断。我们期望一个完善的 NPI 大数据平台能够扭转这个局面，让工程师只需要投入 20% 的时间做数据收集，而能够留下 80% 的时间做深入的数据分析与解析。因此，NPI 制造大数据平台确实有其必要性，并且

非常适用于各个制造行业在新产品开发时的业务需求。

NPI 制造大数据平台主要解决以下两个业务痛点：

1. 数据收集整理耗时且复杂：

由于分析所需的数据横跨五个以上的孤岛系统、且涉及不同部门，工程师需要与不同部门进行沟通，并且切换不同系统以取得数据；而取得的数据格式不尽相同，甚至有些数据为非电子文件，工程师需手动一一将各个脱机数据进行输入、整合、清洗、整理成自己分析所需的格式，如此一来，仅一个问题的数据收集就花费 4 个小时以上了。

2. 涉及信息太多：

数据包含产品各阶段的生产数据（产品测试数据、产线组装数据、进料检验数据、关键物料数据、供货商数据等），且各类数据量都很大（例如工站的监测项就有上千个，甚至上万个）；手动整理数据容易发生错误，且不易察觉错误点；再者，各分析软件有处理数据量的上限，使用分析软件进行实时数据增加、删除皆属不易，如此一来，真是耗时耗人力。

2、 解决方案

本案例主要采集来自不同系统的数据，并通过数据整合处理运算及应用模块（DIF&SMC）进行数据集成与处理并建立分析 workflow 模块、数据查询模块以及知识库检索模块（KM）作为数据建模与分析层来分别提供通用性分析应用、FA 分析平台决策及知识库驱动分析应用。实施概况如下。

1. 数据来源

（1）产品测试数据

该数据为产品在各工站的检测数据，来源包含四个异质系统。

（2）产线组装数据

该数据为产品在工站中的组装流水线信息，来源包含两个异质系统。

(3) 进料检验数据

该数据为产品的物料在进货时的检验数值，来源包含两个异质系统。

(4) 关键物料数据

该数据为产品上所使用关键物料的相关数据，来源包含两个异质系统。

(5) 产品组合数据

该数据为产品的各模组元件相关数据，来源为一个系统。

(6) 关键尺寸数据

该数据为产品的尺寸相关数据，来源为一个系统。

2. 技术方案

(1) NPI 大数据平台整体架构

此平台以基础数据为基底，通过 SMC(Spark, Mesos, Cassandra)三套开源大数据产品将关联性数据进行整合，再由 DIF(Data Integration Framework) 架构进行数据处理运算提供多维度整合性数据至 DSP(Data Service Provider) API、分析工作流等服务，系统应用层利用多维度数据进行通用性分析、FA 分析及查询服务。如图 1 所示，分述如下：

数据采集层：数据采集层应考虑应业务需求如何有效且完整的取得异质系统的数据。因采集的数据来源有内部系统、外部系统、及非系统化的本地文件等，所以需要根据各异质系统进行数据交换，分别利用系统接口、网络爬虫转换文件格式储存及电子邮件方式等交换取得各数据。

数据处理层：此层级运行包含 SMC、DIF 及 KM。数据处理层主要考虑为各异质系统数据进行关联整合，因此必须着重于数据清洗、数据理解及数据的关联，才能够提供分析层有效的模型处理。

- a) KM 主要接收各类型文档进行分类储存、解析内文、并利用 Elastic Search(一套支持全文检索的开源项目)进行建立索引库及支持文件内全文检索服务。

b) SMC 及 DIF 另述于后。

数据分析层：数据分析层包含三个项目，分别为 DSP API、分析 workflow、KM API。分析层构建 API 主要是为了提供分析模型并有效率地产生关联性数据的业务需求而构建分析 workflow。

a) DSP API：通过 API 方式将产品关联性数据提供给 FA 分析平台。

b) 分析 workflow：用户自定义需求数据字段通过 stored procedure 定期向 DIF、SMC 索取关联数据，并整合成多维度大表数据，提供数据应用层分析功能使用。多维度大表在本项目中是一个非常重要的核心功能，透过多维度大表将各种数据源依据用户单位本身业务上的需求，进行数据的关联性分析与整合，所涵盖的数据属性(Attribute)数量从上百个多至上千个字段，满足用户单位在海量数据信息中，挖掘对于业务问题能够使用的关键小数据，进行问题成因分析。

c) KM API：数据应用层透过 API 进行档案增、删、修行为，且进行字段式及全文检索式的档案搜查。

数据应用层：NPI 大数据系统包含三大应用模块，分别为 FA 分析平台、通用性分析、以及知识库搜寻模块，分述如下。应用层实现用户在单一平台中快速取得需求数据，可于八成时间内处理日常作业（FA 分析平台），并可利用其余二成时间进阶分析数据（通用性分析）。再者，知识库可以满足工程师之间的技术传承及新人训练，帮助自主学习成长，并减少工程师的沟通教学时间。此应用让工程师的时间得到最有效的利用。

a) FA 分析平台：工程师在进行 FA（Failure analysis）时，通过 API 实时取得产品组装、测试结果、物料资讯等更多相关数据，减少工程师四处搜集数据，加速问题解决，可取得更多的时间进行更多、更进一步的分析作业。

b) 通用性分析：工程师于 FA 工作之时或之余，利用通用性分析 workflow 提供的多维度表分析批量数据，通过七大分析工具（散点图、箱体图、常态分布图、直方图、线图、良率报表、测项名称比对）洞察数据潜

在的问题；并可针对 golden case 建立定期报表，由系统自动根据设定条件定期检验潜在问题，工程师可侦测问题并实时解决问题。

- c) 知识库搜寻模块：工程师在解决问题时可通过搜索模块来取得相似问题查看，无需一一向他人询问，即可以学习前人经验，以达到在线学习及经验的传承。

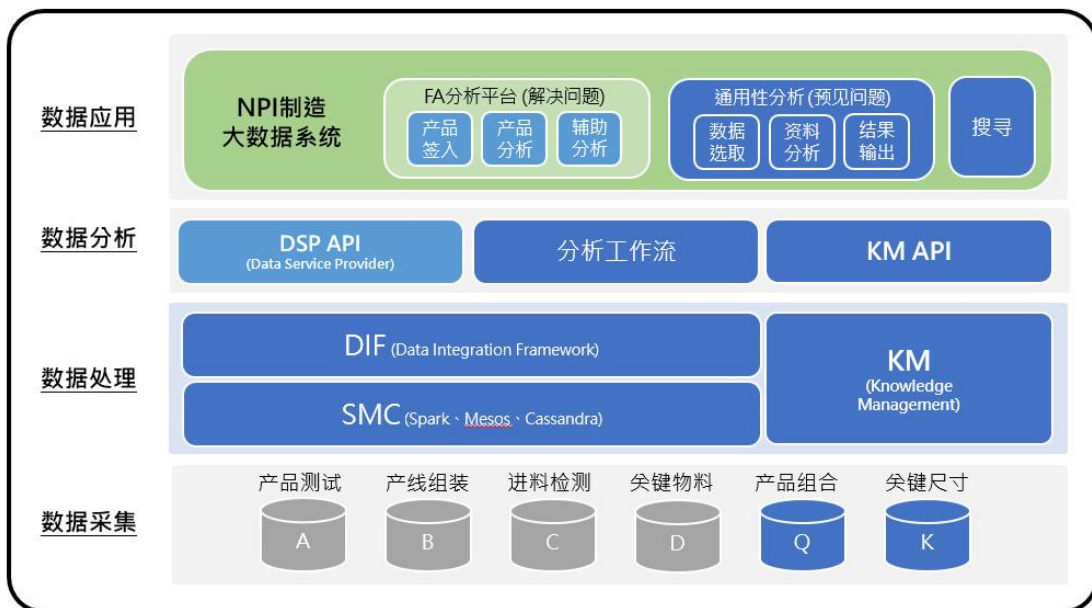
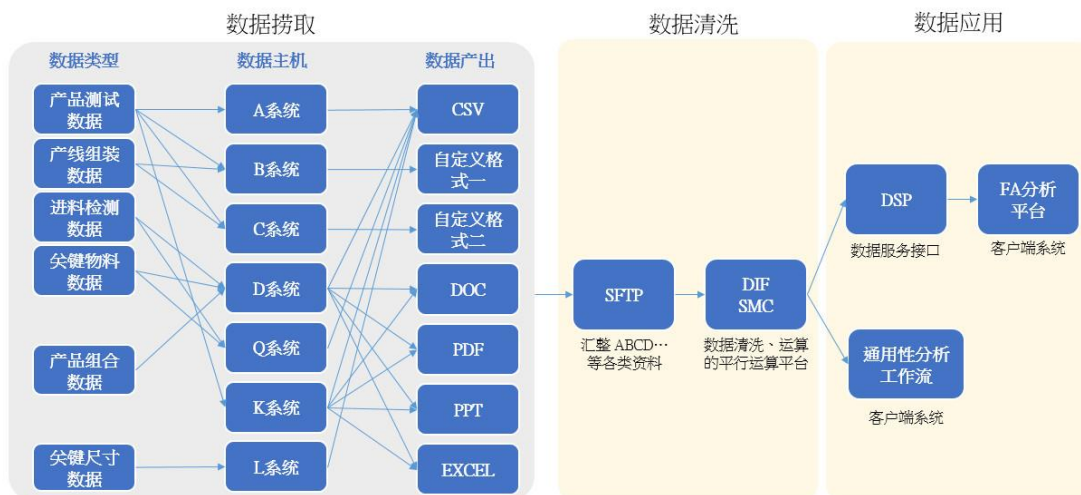


图1 平台整体架构图

(2) 数据流

本案并没有通过传感器直接读取产品参数，因为本项目所收集的数据主要以工程导入及测试记录为主，故可以分成三大类: 1.通过生产系统的接口获取数据，2. 制造单位定时发送 email 提供物料组合配方信息 3. 利用网络爬虫扫描系统接口查询报表转换数值数据等三个主要方式。收集的数据类型包括了：产品测试及组合、产线组装、进料检验、关键物料及关键尺寸数据等。构建数据整合与应用模块，将收集到的数据进行清洗、整理等工作，并依需求产制相关多维度大表提供应用，以达到基础数据完整的目标。



(3) 基础数据整合架构

DIF 主要包含三个模块：数据整合、数据处理运算、及信息呈现。

数据整合（Collection Modules、Transfer Modules）方式主要透过实现一个真实案例所需的数据来源，为验证案例并进行调整。

数据运算（Analysis Modules）则是将真实案例所需要的参考信息，预先整理（计算、统计、多系统串连）汇整成一信息整合包，形成一个数据分析处理流程。

信息呈现（新应用模块）提供一个符合用户需求的操作接口及相关应用工具，并整合于原系统中，产生新的应用模块。

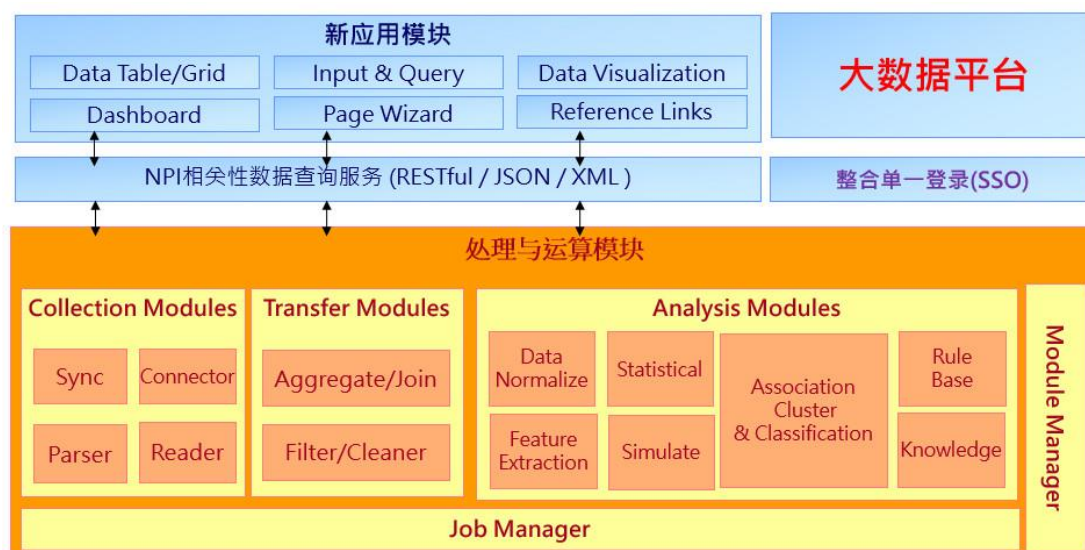


图3 DIF 架构图

3. 项目其他亮点

在大数据平台中特别增加两种矩阵式分析模型：工站良品率分析矩阵图及物料组合优化矩阵图。工站良品率分析矩阵图：主要查看工站测项在不同时间段下的趋势表现。物料组合优化矩阵图：主要不同物料在不同供货商组合之下的趋势表现。用户可通过工站良品率分析矩阵图得知表现最差的测项，并通过物料组合优化矩阵分析图来针对该测项取得最佳物料的供货商组合，以达到良品率的提升。

3、 实施效果与推广意义

1. 时间短：数据整合收集所需时间大幅减少，由以前耗时 2~4 小时，提升至数分钟以内。

2. 数据广：整个 NPI 周期大约会生产 5 万个产品，而 NPI 产线大约 150 个相关产品测试数据完整收集，涵盖 400 多个关键物料及平均 2000 个测试项目，特定工站更高达 4 万个测试项目。NPI 大数据平台每日数据吞吐量大约 22 万笔原始数据档案。

3. 效率快：工程师将原来数据收集的时间专注用来解决问题，并可利用剩余时间进一步分析预见问题。扭转过去 80/20 法则，让工程师只用 20% 时间收集数据整理数据，便能充分投入 80% 时间专注在问题解决上。

4. 传承快：前辈工程师将个人的经验时时累积于平台上，新人工程师可随时事半功倍地学习前人的知识。

4、 案例亮点

利用 NPI 大数据提升 FA 工程师的数据分析能力。

分析工程师依经验及型态可区分为五种类型。经验值较一般的分析工程师可在 FA 分析平台上解决问题并同时学习前人的经验；经验值较高的分析工程师则在通用性分析中预见问题，以加速预防问题；有经验及具备统计软件使用经验的分析工程师可利用多维度大表分析解决客户问题；具备研发能力的分析工程师可

运用建模软件进行分析模型研究；具备算法能力的分析工程师则通过程序开发最新且先进的模型研究。每个步骤都是可累积经验并进行学习成长的。

此平台不仅可解决问题、预见问题、累积知识、经验传承，并通过多维度数据的串联供不同阶段工程师进行开发、研究以及回馈，使得工程师与系统共同成长。

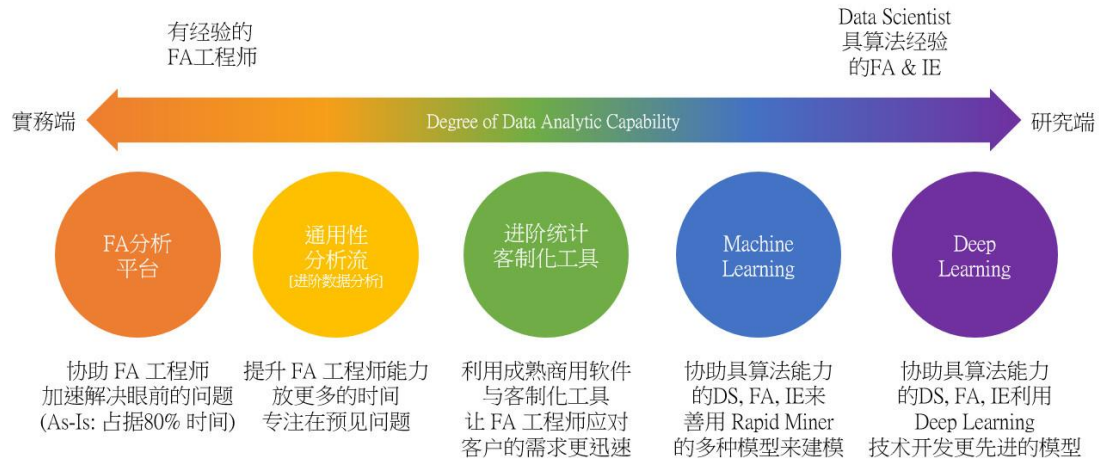


图4 数据分析阶段性应用

二、 中国联通工业大数据采集分析平台

1、 案例背景与业务痛点

为贯彻落实《中华人民共和国国民经济和社会发展第十三个五年规划纲要》和《促进大数据发展行动纲要》，加快实施国家大数据战略，推动大数据产业健康快速发展，工信部编制了《大数据产业发展规划（2016—2020年）》。中国联通大数据业务开展的非常早，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。中国联通大数据平台可以实现对 IT 生产系统、业务平台、通信网络、外部互联网、合作伙伴 5 大类数据采集，覆盖客户数达到 4 亿，日处理数据超 4000 亿条，高价值结果数据沉淀超过 17PB。大数据水平处于国内行业领先地位。工业行业是中国联通产业互联网战略的重点方向，而在当前的制造工业实施过程当中，很多的机床由于历史原因，无法通过数据总线采集上报机床的关键运行数

据，企业在生产过程当中对产品的品质控制、良品率的提升无法开展。因此在国内大部分制造企业不更新换代设备的基础上，如何采集、分析生产过程中的各种关键数据，是目前实施工业互联网战略的重要难题。

1. 项目简介及目标

中国联通工业大数据采集分析平台针对生产车间内的机床、环境等各种可能影响产品质量和成品率的因素进行了采集，通过大数据的方式实现对生产过程的优化、机床配置数据的改进等，以提高企业的生产制造良品率。

2. 项目目标

平台的目的是通过工业级的采集传感器进行第一手关键生产数据的采集，通过无线物联网的方式回传至大数据平台，以大数据分析的结果反向回馈到生产制造的设备配置、使用、工艺改进等一系列过程中，以便企业能有效的提高产品的质量和成品率。

2、解决方案

1. 数据来源

大数据平台数据来源主要包含三大类：

（1）生产传感器数据

主要包含了机床生产制造过程环节数据，主要有类似应力、表面温度、传输压力、传输流量等数据。目前以单车间 1000 个采集传感器来计算，平均每 20 秒上报一次数据，数据包大小以 200KB 为平均量，那么单个车间每天将有 824GB 的数据量产生。

（2）车间环境传感器数据

环境传感器采集工厂内部的各种环境数据，提供对机床设备运行和设备使用寿命，工厂内工作人员的健康状态的影响评估。采用 NB-IoT 物联网络进行数据实时采集传输。

（3）现场总线上报数据

主要针对工厂内机床设备的对应配置参数，可以提供总线数据开放的部分进行采集。

（4）生产系统数据

对接生产系统的排产计划，产品的技术检验技术要求等数据

2. 技术方案

中国联通大数据采集分析平台主要集成工厂数据采集、大数据分析平台能力，具备支持各类合法的通讯芯片和模组、传感器、连接管理平台、计量器等仪器仪表设备接入，加工并处理数据资源，支持多种应用的分发和部署，实现资源、组件能力的快速部署、管理，解决工业互联网设备的数据接入问题，降低使用的技术门槛及运营成本，提升开发效率。在系统层面上，平台具有向上与微服务和能力化应用对接，提供基于工厂生产数据的服务，向下具备收集基于工厂传感器数据，兼容各种物联网设备，支持数据的实时和非实时传输。

在能力层面上，平台应具备数据采集能力及大数据能力，涉及数据存储、清洗过滤、分析挖掘、接口封装等过程，实现各类资源的统一调度以及各种能力的快速部署。

物联网大数据平台类框架如图 1 所示。

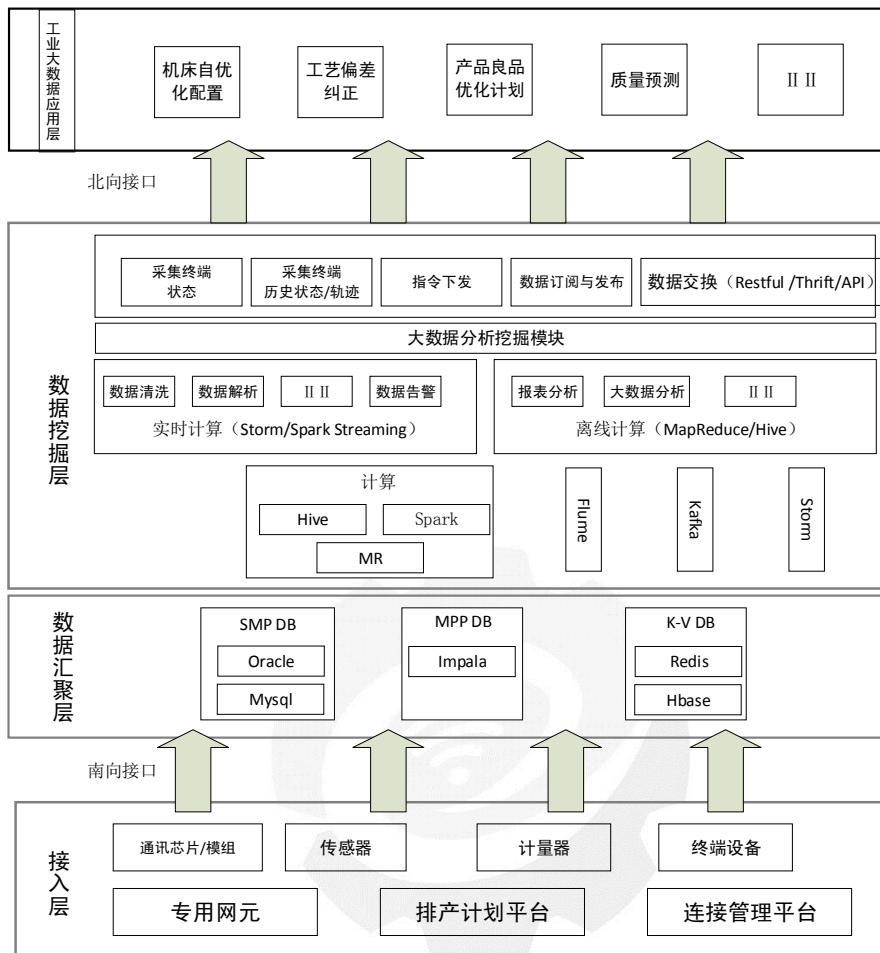


图1 物联网大数据平台框架

大数据平台提供工厂内大数据应用所要求的各个性层次、规模层次的数据存储、数据加工作业、数据分析挖掘、数据安全审计等功能，保障用户数据安全和 服务稳定，并提供一系列工具协助用户开发调试、监控性能和优化任务执行的大 大数据能力平台。工业大数据平台架构主要包括数据接入层、数据汇聚层、数据挖 掘层、数据应用层、以及相关的数据接口等部分。

(1) 数据接入层和数据汇聚层：解决数据获取和管理

数据接入层主要功能是通过通讯芯片、模组、传感器、连接管理平台、计量 器、以及终端设备等仪器仪表获取的环境、资产、运营状态等数据信息，是数据 汇聚、存储、加工等操作的前提条件。

数据汇聚层主要功能是将传感器终端设备采集的数据信息，按照相关要求进 行清洗、入库、存储操作。

（2）数据挖掘层：解决物联网数据实时和非实时处理

数据挖掘层主要功能指对经过清洗加工后的数据进行数据分析挖掘操作，按照实际需求，通过数学建模等手段，充分挖掘采集到的数据中所包含的信息。

（3）数据应用层：解决工厂内大数据应用模式问题

数据应用层的主要功能是将平台功能与实际行业相结合，提供基于真实的应用案例的若干典型应用模型框架供用户直接使用，也提供复杂应用定制的服务。

3. 项目其他亮点

工业级的传感器数据采集解决方案为工厂内特殊环境的数据采集提供可靠的解决方案，解决目前常见环境的传感器无法应用与工业生产环境的问题。

3、实施效果与推广意义

客户企业在应用本平台解决方案后，有效提高了生产设备数据采集效率，并对设备的维护效率有很大提升。同时通过大数据的采集和分析提高产品的良品率，达到了降本增效的目标。保守估计，项目解决方案全面部署后，每年将为客户节省约 3000 万成本。

4、案例亮点

采用外部传感器数据采集方式完成非开放总线数据生产设备和非数据化生产设备的数据采集工作，将无法联网无法数据化的生产过程进行数据化，在线化管理，为工厂生产过程中的工艺改进，良品率提升提供了数据基础。

三、 华为云 EI 企业智能打造智能化九州通

九州通是一家以药品、医疗器械、生物制品、保健品等产品批发、零售连锁、药品生产与研发及有关增值服务为核心业务的大型企业集团。目前公司针对 B 端客户有 15 万多家，C 端客户已达 1050 万以上。旗下医药流通公司达 130 多家，覆盖了中国整个行政区域 95% 以上，在新疆、西藏都有子公司，并在未来三年，

公司计划拓展到 300 到 400 家的规模。2016 年，九州通的营业额达 615 亿，不过九州通的野心远远不止是数字的增长，除了要在 2019 年营业收入达到 1016 亿之外，还要实现一个智能化的九州通。

同时，九州通认为自己面临两大挑战：第一是持续创新，只有创新，才能从容面对市场竞争，有助于构建企业自身核心竞争力；第二是对市场变化和企业自身业务发展需求的快速响应，传统靠业务驱动 IT 的发展方式是不够的，九州通认为更应该是战略驱动 IT，要更多关心企业将来发展战略方向是什么，所以未来九州通的企业定位是要做业务创新的推动者。而 IT 或技术本身就是第一生产力，在业务同质化的今天，更多的是强调 IT 的创新，现在 IT 创新就是要依靠云技术。这时，九州通和华为云走到了一起。

1、案例背景与业务痛点

现如今，在阿里巴巴，京东等企业的带领下，物流行业发展迅速，先后实现了无人仓，机器人拣货等快速物流作业方式，而九州通依旧沿用人工拣货方式。导致在企业业务快速发展的同时，物流却止步不前，越来越无法满足业务的需求。

1. 拣选路径优化

在医药行业中，发货时比较零碎并且不集中的情况比较严重，九州通下属的二级公司 1 个拣货员一天拣货平均明细为 1,000 条，步数 30,000 步（约为 20 公里），导致拣货员超负荷作业，并且效率有明显的不足，急需优化拣货路径，提升效能。

2. 派车路径规化

目前九州通的订单量在医药行业是比较大的，每天的订单量动辄上千，派送车辆动辄几十辆。在调度每辆车的派送订单和送货路线时，调度员依据经验选择订单集合和派送点集合。结果是，车辆轨迹在地图上十分混乱，整体成本居高不下。

3. 智能装车

目前九州通的配送模式为当天配送，但是在配送过程中对于客户的距离远近及配送顺序都是凭借配送司机经验来进行选择性的配送，在这个过程中无法准确的判断哪些客户群体距离比较近、配送比较快；哪些客户距离比较远、需要最后进行配送等，并且在配送过程中对于小客户很多时候无法很好的支持配送到位，对于客户方面，客户也无法具体知道自己的药品当前还需要多久送到。因此希望通过智能装车与配送达到根据客户距离远近及配送紧急程度等要求，来计算出最优配送线路与到达每个客户的时间点，以便于配送员配送且能让客户知晓当前配送状态以及何时能到货。同时，针对部分小客户也能达到支持配送的目标，提高小客户满意度与订单数。在装车的过程中，若能告知如何装车，则能够将车辆的容积利用最大化，减少车辆利用率。

4. 筛选热销品种，提升出库效率

在医药行业中，物流仓库的货位零散多变，对于热销品并没有固定货位，也无法根据季度来判断哪些是热销品，因为货位的多变，很多热销品就藏茫茫货位中，拣货员总是要去寻找，耗时耗力。若能通过计算一个月及一个星期的效率来进行预判，哪些品种当前会销售多少，将热销品的货架固定在距离复核台最近的地方，在拣货员作业时，能达到 80% 的时间都是在热销区域进行作业，这样针对作业效率的提升有助于当前整个作业的提升。

5. 识别商品电子监管码

按照国家药监局的要求，每个药品都需要追溯其流向，而电子监管码就是药品追溯的唯一条码，在公司进行物流作业的同时，需要将其录入到系统中，方便追溯，现有模式是采用扫描枪进行扫码复核，一个任务若有 100 个商品，则需要花费 30s 左右，若利用 OCR 技术进行识别扫码，希望可以达到 10s 完成。

6. 发票识别

按业务日常需要，在勾兑上游客户发票的时候，必须要发票和单据明细挂钩，现在的处理是要采购员来对着发票内的明细进行一一勾兑和录入，操作简单，但繁琐耗时间，若能够通过 OCR 技术实现自动勾兑自动录入，则会很大部分的节

省人员时间，提升效率。

7. 销售预测

对各零售平台的数据无法准确预测，指导生产和物流。

8. 大数据平台

集中平台，无法将能力较好的开放出去给其他营业部门使用

2、解决方案

九州通公司基于华为云建设工业领域集成平台，联接企业、供应商和客户，打通业务全流程，打造全面协同的医药产业链。九州通公司希望通过“云和物”实现全联接，做智能化九州通公司。

本项目中主要业务痛点和项目目标为：

(1) 派车路径规划：调度车辆和送货路线时，依据经验选择订单集合和派送点集合。导致车辆轨迹在地图上十分混乱，整体成本居高不下。

(2) 仓储规划：物流仓库的货位零散多变，拣货员总是要去寻找，耗时耗力。

(3) 发票识别：靠人工发票内的明细进行勾兑和录入，繁琐耗时间。

(4) 销售预测：对各零售平台的数据无法准确预测，指导生产和物流。

(5) 大数据平台：集中平台，无法将能力较好的开放出去给其他营业部门使用。

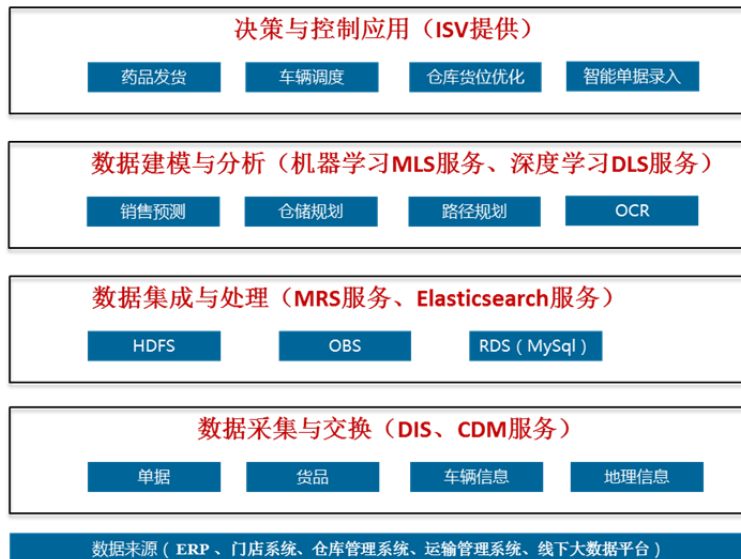


图1 九州通 EI 企业智能解决方案架构图

1. 路径规划方案

华为云智能物流服务通过高效的机器学习和优化算法解决了多个物流场景的优化问题，并同时能满足实际场景中的多项限制条件。路径规划是其中一个场景。

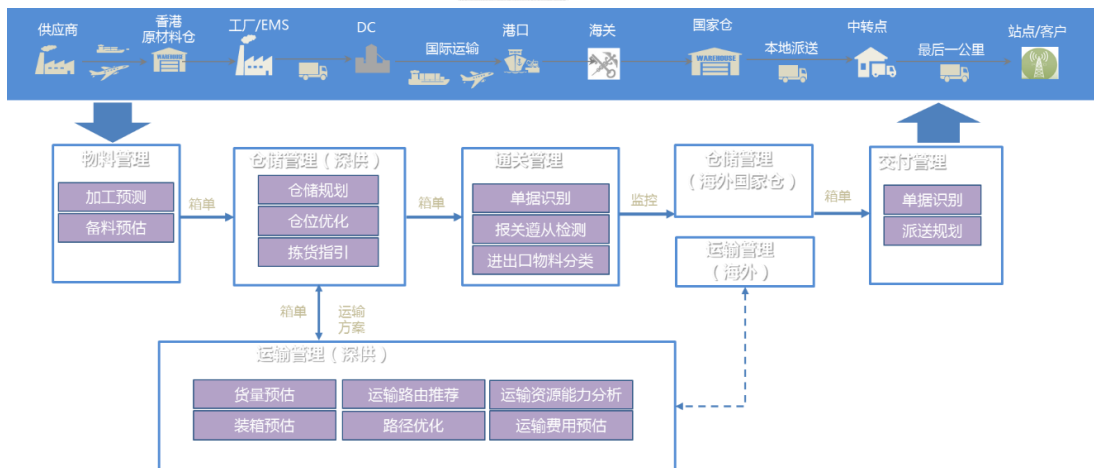


图2 华为智能物流整体解决方案

路径优化关键输入、输出分析详解：

该服务主要解决多订单多地址运输的物流优化，根据订单信息以及地图信息，在满足订单限制和运输工具限制的前提下，基于用户给定的距离信息或地图信息，给出最优的运输策略，降低运输成本。该服务的输入是订单以及运输工具的信息，输出为运输工具的运送策略。

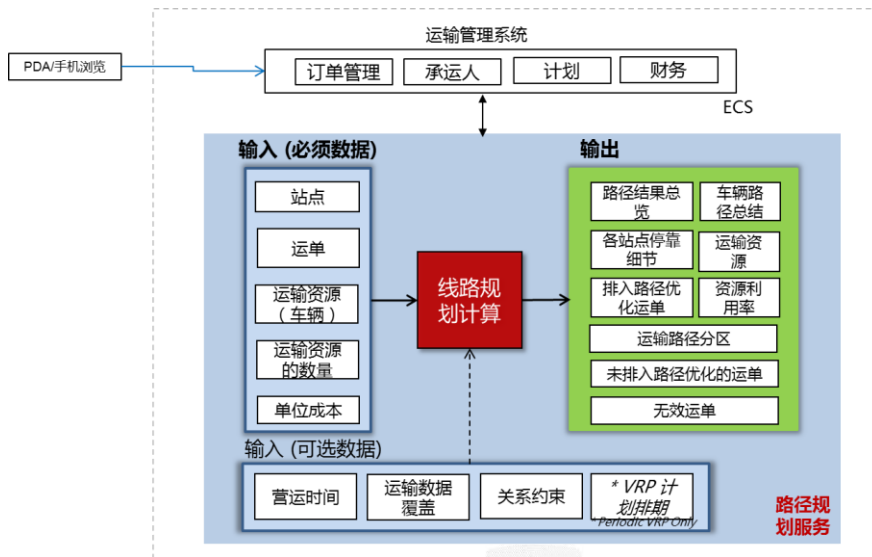


图3 路径优化方案架构设计

方案亮点：

- (1) 提供两套算法（蚁群和遗传），根据实时性要求进行选型：
 高实时性（1分钟内）：蚁群算法，少量迭代能获得次优解；
 低实时性（半小时内）：遗传算法，收敛时间长，能获得更优解；
 训练好的模型，直接发布使用；客户直接远程进行 API 调用即可，使用简单。
- (2) 通过先全局分区域聚类订单再局部优化路径，能达到更加优化的效果。

2. 大数据平台云上迁移方案

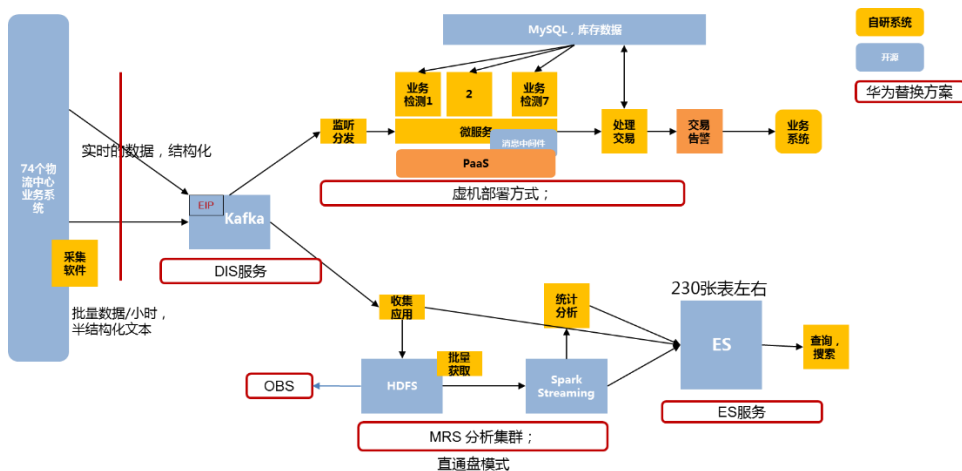


图4 九州通大数据平台云上迁移方案架构

数据来源:

- (1) ERP 系统销售数据
- (2) 门店发票
- (3) 仓库管理系统
- (4) 运输管理系统
- (5) 已有大数据平台

实施方案:

(1) 数据源存储在原始 Oracle、MySQL 等数据库中，数据通过已有的数据采集软件和公有云实时消息 DIS 服务，用于实时增量抽取数据上公有云；

(2) 对于生产业务系统，搬迁上云后，采用华为 CCE 云容器技术，方便应用的快速迁移和扩、减容；同时传统 MySQL 数据库可以用公有云服务 RDS 来替换；

(3) 需要实时处理的数据一部分会通过 SparkStreaming 做基于窗口的实时统计分析，一部分会基于 Elasticsearch 构建分词检索，用于后续日志查询；

(4) 所有原始数据无需存储在本地 HDFS 盘中，只需存储在 OBS 对象存储中，通过 MRS 集群及时分析，结果写回 OBS，极大降低了成本。

方案亮点:

(1) 大数据技术栈的全集，全面兼容开源大数据接口，方便线下大数据快速上云

(2) 集群全托管模式，自动集群创建、扩减容管理

(3) 按需使用，存储、计算分离，节省 50% 以上成本

(4) 入门简单，界面友好，简单点击几个按钮就可完成

(5) 无需购买安装服务器、手工部署和调优 Hadoop、Spark

3. 销售预测分析方案

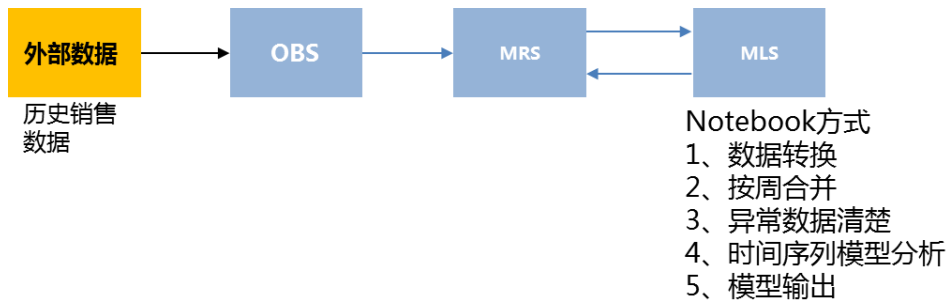


图5 销售预测分析流程图

通过 2~3 年的每天的的历史销售数据，预测下周或下月的销售额度，项目中涉及近 3 年各类千余种药品的大量销售数据，几百万条销售记录。

(1) 异常数据识别分析：原始数据中有客户是累计几天才上报一次，导致某天的销售额极高，影响正常数据的预测。通过 PCA 统计分析模型，提前将异常数据过滤掉，保障数据的精确性。

(2) 数据预测周期分析：一般周六、周日销售额都比较低，如果按天维度进行预测，会出现很大的波动；后面统计认为，按周分析是最佳度量。

(3) 预测算法：采用时间序列模型分析，能够对长期趋势、季节变动、循环变动等进行建模预测。

(4) 模型固化：模型固化成 Notebook 模板，需要预测时只需要更新原始数据，直接 Run 一遍即可。

方案亮点：

- (1) 全流程平台化操作，支持 Workflow 和 Notebook 拖拉拽方式建模
- (2) 能够固化算法模型，后面数据定时导入预测，模型自动修正预测
- (3) 算法模型可方便共享发布，支持标准 PMML 模型文件和 jar 文件方式
- (4) 训练的模型、评估结果每个阶段均可可视化支持，提升模型的可解释性

4. OCR 增值税发票单据识别方案

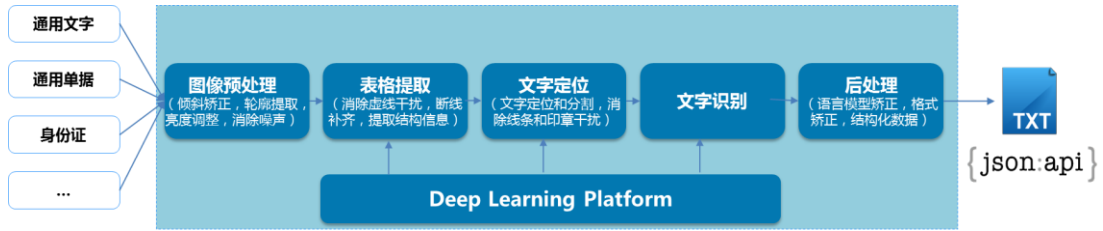


图6 OCR 单据识别架构图

通过 3 个多月，针对近 5,000 张原始单据以及近 10,000 张其他单据的训练，识别准确率提升到 98% 以上；并集成常用药品名称语料库，大幅提升识别准确率；内置了表格自动识别算法、文字分离模型、小数点、字符等后处理增强算法。

方案亮点：

能较好的处理如下场景：

- (1) 图像模糊、扭曲、倾斜、噪声
- (2) 中英文混合、小数点、特殊字符
- (3) 盖章、错行支持
- (4) 表单格式多样支持

3、实施效果与推广意义

1. 派车路径规化：

经过使用华为云仓库拣零路径规划服务后，拣货员的拣货路径明显规律许多，总体拣选步行里程减少到 20,000 步出头，效果提升近 30%。

经过使用华为云路径规划服务的优化后，车辆配送效率显著提高，在严格满足运行时间的限制条件下，同一批订单派送所需车辆有效减少，平均每辆车的派送能力提升近 5 倍。

在采用华为云智能装车服务后，由于送货批次在装车时已经考虑，所以在装卸货时的效率提升；同时，车辆的装载率也提升了 10%，结合较优的路径规划结

果，整体配送成本大幅降低。

2. 仓储规划：

通过使用华为云仓库入库储位优化，计算出药品的整体重要程度，然后将重要性高的药品摆放到出库成本低的储位，大大降低了拣货员的作业成本。

3. 发票识别：

通过使用华为 OCR 识别服务进行扫码识别，药品复核时间从 30s 降低到 10s 内，大大提升业务效率。

通过使用华为 OCR 识别服务进行发票识别，实现信息自动勾兑录入，大大提升采购员的效率，节省人力成本。

4. 销售预测：

更有效的指导生产，同时能实时更新预测模型。

5. 云上大数据平台：

让开发和平台能力共享更加的便捷，降低九州通公司自身运维成本。

4、案例亮点

项目的先进性与创新点在于充分利用华为云所提供的企业智能服务（EI），用人工智能、大数据等技术，同时采取公有云和线下结合的方式，精准找到企业可快速改进点，迅速落地，帮助九州通公司改善物流路线规划、仓储规划、发票识别、销售预测等业务需要，一步步朝着智能的九州通公司迈进。

（1）智能物流服务：提供灵活可定制的 API，能满足客户不同的个性化需求。

（2）OCR 服务：通过高效的深度学习和图像处理算法，实现了多种类型单据的文字内容识别，满足客户信息自动化录入需求，大大降低客户人力成本。算法的性能和精度均领先同行。

（3）机器学习平台：通过 Notebook 可编程方式，固化销售预测模板。

（4）大数据云平台：EI 大数据平台迁移线下系统，更利于全国接入提供服务。

四、 潍柴工业大数据平台

1、 案例背景与业务痛点

潍柴工业大数据平台是工信部“智能制造”专项项目大数据平台的重要组件，同时也是智能制造必要的平台和支撑。智能制造对 IT 技术提出更高要求，急需工业大数据平台做支撑。

多个智能制造项目均需分别建立大数据平台。平台重复建设耗费精力，项目数据之间无法互联互通，制约数据关联及价值挖掘的潜力前提下，需要全新的技术能力与架构提供更强大的数据存储、治理、计算和调度能力。同时潍柴搭建统一的大数据运营管理平台需求明确，建设企业统一的大数据运营管理平台，做数据积累，提升数据价值密度，为后期数据挖掘提供平台支撑。

1. 项目背景及业务痛点

潍柴作为中国传统制造业的杰出代表，需要进一步适应工业发展潮流，占领工业 4.0 时代的科技制高点，引领行业发展。潍柴传统制造业转型升级的方向就是迈向智能制造。在“中国制造 2025”框架下，智能工厂的标准体系是非常重要的一个环节。潍柴当下正在积极实施工业 4.0 的建设，搭建工业大数据平台，从底层实现数据采集和互联互通，从数据层面洞察制造信息，进而打通从研发到生产管理的整个流程，为企业的生产和发展提供新的模式和思路。

通过电信对潍柴大数据团队培养，可获得大数据系统建设能力。本项目将获得与中国电信共有的大数据服务能力，通过二次定制开发，电信可支撑潍柴对外提供相关服务，将能力有效商品化。在支撑潍柴内部业务同时，可适时商品化成为一家专业的工业大数据服务提供商，填补国内工业大数据服务空白，增加工业制造领域影响力。为我国智能制造树立标杆，将产生广泛的社会价值。

生产排程管理作为智能制造的一个重要环节，本项目通过融合生产排程与工业大数据技术，可以提供精准化、智能化的生产排程能力。

2. 项目简介

潍柴大数据平台围绕“互联网+工业制造”建立一套互联网+制造融合创新模式平台，平台在工业应用云、公有数据云、私有数据云建立智能应用平台。通过电信的云基础平台结合工业 PON、LTE 专网实现工业智能网关的感知并接入生产相关制造数据，形成潍柴工业大数据云。

应用层主要实现企业内工业的微应用连接，包括智能制造类、产品设计类等。企业外工业微应用连接包括协同化设计、预知性维护、服务转化型等等。以工业连接为起点，实现用户中心与生态协同的应用连接，以数据数字驱动为导向建立标准体系平台。

3. 项目目标

打造自主工业大数据平台，建立引领行业的工业大数据标准。

提升企业在行业领域的品质竞争力和成本竞争力，依托工业大数据平台，在数据算法模型研究应用领域形成核心技术能力。

(1) 业务优化目标

利用大数据分析技术对生产节拍数据进行洞察，对各种影响生产节拍的因素进行特征提取，建立相关因素的关联规则，结合发动机订货信息，形成工业 4.0 架构下的精益排程，进而对潜在的问题进行分析预警。

(2) 系统建设目标

本次项目建设包括产节拍数据的自动化采集、工业大数据平台和生产排程应用三部分。

2、解决方案

大数据平台建立基础大数据平台及系统优化，搭建大数据存储相关计算平台，大数据分布式系统。建设全文检索平台，开放全文检索数据访问接口及集成大数据存储接口。建设数据接入层服务框架，基于 HTTP 和 FTP 相关数据接入服务，基于 Agent 的海量数据实时接入及传输服务。

1. 数据来源

(1) 潍柴测试机床拧紧机相关数据

拧紧机在配合拧紧程序会产生拧紧过程并在拧紧过程中产生数据，通过拧紧机自带程序对采集到的数据进行汇总并上传至本地存储服务器完成原始数据采集。

(2) 潍柴测试机床节拍器相关数据

生产节拍数据需要采集全部产线的工位节拍数据，数据采集分为两种情况：第一种是工位节拍数据已经保存到企信部历史数据库中，可以手动或定时将数据从历史数据库中导出，并上传到工业大数据平台；第二种是新增的图像采集点，直接把图像识别的数据上传到工业大数据平台。

2. 技术方案

(1) 大数据能力平台架构



图1 潍柴工业连接+大数据平台的总体架构

■ 基于 FTP 的文件型数据采集服务

大数据平台通过 FTP 的 get 服务去其他数据源获取数据，并存储到大数据平

台。

■ 基于 Agent 的海量数据实时接入及传输服务

在数据源初部署 Flume agent，通过 Flume 获取数据源数据，并存储到大数据平台。

■ 关系型数据库数据接入服务

提供关系型数据库和 Hadoop 平台之间数据相互转移的功能。

■ 基于 Http 的数据接入服务

数据源通过 HTTP 服务发送数据到大数据平台。

■ 全文检索和大数据存储平台接口集成

实现 ES 中索引数据和 HBase、HDFS 中原始数据的关联的建立，对外提供一致的索引和原始数据的访问接口。

■ 关系型数据库数据接入服务高可靠性组件开发

基于开源 Sqoop 数据抽取框架高可靠性开发，避免数据丢失。

(2) 大数据平台功能架构

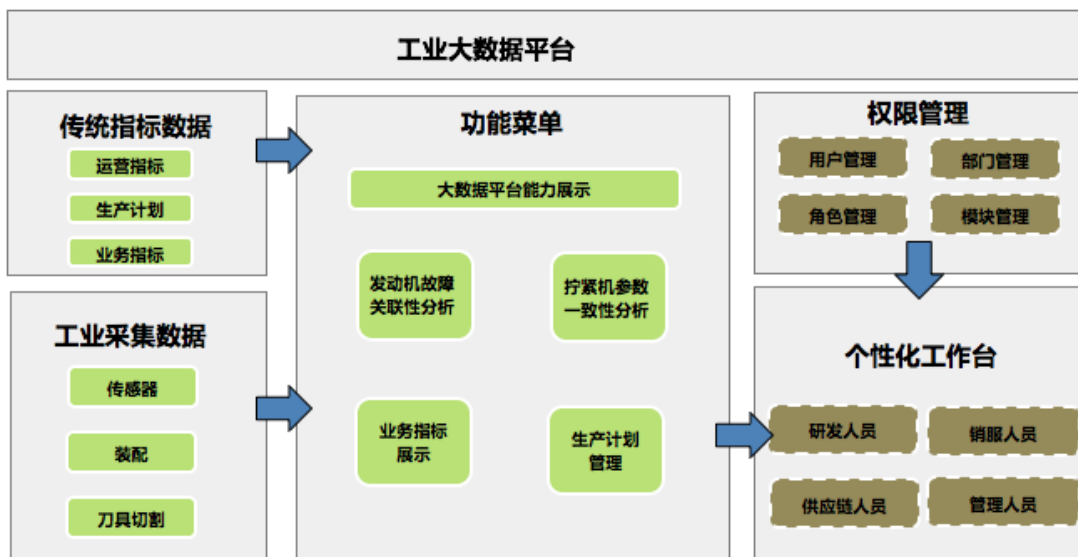


图2 工业连接+大数据平台功能架构

通过数据传输、大数据本身数据处理能力，来展示大数据平台的能力（包括CPU/内存、存储能力，数据传输能力等）。点击传输的数据，可以从时间维度清晰了解数据产生的趋势和工厂设备运转的关系。

3、实施效果与推广意义

定量指标：经济效益（直接收益）、工作质量（出错率、废品率等）、工作时间（标准作业时间）、工作强度（操作简化、步骤减少）、人力节省、风险控制（风险概率减少）、损失控制（损失减少）。

定性指标：流程优化，管理效益（内部客户满意度、人才培养、知识转移），协同效益（对现有平台的充分利用集成），社会效益（安全环保、节能减排）。

4、案例亮点

大数据平台通过数据采集后对螺栓拧紧的拧紧方式、参数进行分析，通过《基于过程数据的螺栓紧固工艺研究建议书》对螺栓数据进行分析。大致的分析结果从扭矩、角度、斜率等结果来判断螺栓一致性，例如在分析某螺栓数据形成过程结果如下：

(1) 采集各项系数分析对机器螺栓扭矩对角度分布；

(2) 大部分螺栓拧紧机分为两个步骤完成拧紧流程，同时可以查看每个螺栓拧紧角度、扭矩、以及第二阶段的斜率进行查询，也可以进行特定时间范围内同样的拧紧数据进行对比分析。算法分析模型采用线性回归，利用称为线性回归方程的最小平方差函数对一个或多个自变量和因变量之间关系进行建模的一种回归分析，这种函数是一个或多个称为回归系数的模型参数的线性组合。

对曲线进行清洗后，分段进行线性回归，可以得出每颗螺栓，在每一段（阶段1和阶段2）的回归线斜率（T/A）然后对相同种类多颗螺栓进行汇总统计，通过统计图的特征，判断螺栓的一致性是否良好（曲线越瘦高说明螺栓一致性越好）

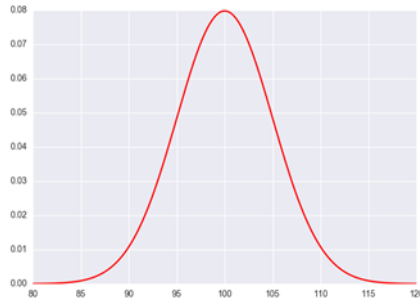


图3 螺栓斜率统计图

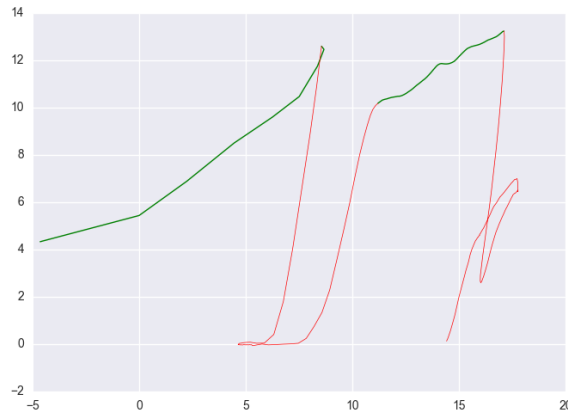


图4 螺栓清洗过程趋势

通过一定的数据清洗算法将与拧紧过程不相关的数据去除如上图红色部分。

利用大数据分析技术对生产节拍数据进行洞察，对各种影响生产节拍的的因素进行特征提取，建立相关因素的关联规则，结合发动机订货信息，形成工业 4.0 架构下的精益排程，进而对潜在的问题进行分析预警。

五、 中国移动 OneNET 助力工业资产管理

1、 案例背景与业务痛点

制版企业生产设备资产管理项目是中国移动物联网开放平台 OneNET 为制版企业的核心制造机床设备--电雕机以及其他制造设备提供的工业物联网设备资产管理解决方案。该项目提供了设备定位、生产监控、产能分析、异常告警、

耗材预测、故障诊断和资产管理等一系列功能，充分利用了中国移动在无线蜂窝网络（NB-IoT/LTE）、物联网平台设备接入和大数据分析处理等方面的技术优势，通过资产建模、数字孪生等技术方案，有效地实现了设备的数字化，同时提供了一个可复制、可持续发展的工业模式，推动了该工厂在工业互联网方面的创新发展。通过对其数据进行工业大数据分析，还可以及时发现管理上的漏洞，有效避免资源浪费，进一步提升资产利用效率，精益化生产，为企业的生产经营管理提供更好的支撑。

1. 项目背景

客户企业是全球规模领先的专业制版企业，拥有近千台业界高端的电子雕刻机、激光雕刻机，在全球拥有近百家工厂。公司规模庞大带来了管理和生产上的诸多问题：

问题一：如何更有效的管理子公司核心生产设备资产？

分散在各地的大量子公司、工厂使得集团总公司以及各地总公司的管理费用急剧提高，尤其是跨国管理海外资产，企业获取真实有效且实时的信息需要投入大量人力财力，如何更高效的实时在线监控高价值生产设备的产能，进行数据统计分析，从而提升核心资产利用效率成为企业制约企业发展所面临的巨大挑战。

问题二：如何更好的管理耗材库存？

核心资产在生产中往往需要消耗大量的配件耗材来配合生产，如何对耗材库存进行优化，降低耗材闲置周期，减少耗材虚报，从而提高现金流降低成本，是企业进一步降本增效的关键点。

问题三：如何提前发现生产中的问题？

生产环节中出现问题往往要等产生了严重的后果之后才能发现，如何及时准确的发现异常来大大降低企业的损失，如何减少人力投入来节省质量成本，以及如何实现预防性维护和故障诊断都是企业迫切希望解决的问题。

问题四：如何解决私自生产，虚报产量的问题？

在生产管理方面，特别是在公司庞大，工厂数量多，工厂 IT 系统不完备的情

况下，各地工厂可能存在私自生产，虚报产量，篡改生产数据等问题。企业集团亟需未经污染的真实数据作为凭证手段进行企业生产管理。

2. 项目简介

制版企业生产设备资产监控项目是中国移动物联网有限公司为制版企业的核心制造设备--电雕机提供的工业物联网端到端的解决方案以及方案实现。中国移动物联网公司利用其在物联网接入、大数据分析、应用使能等方面的技术优势帮助制版企业成功实现设备资产云端数字化，3D 虚拟巡检、产能监控和资产管理等功能，进而推动了该产业在工业互联网方面的创新发展，特别是利用大数据分析和数字孪生技术初步实现了智慧工厂的部分功能。

3. 项目目标

该项目核心是实现制版企业核心制造设备的云端数字化、产能监控和维护性优化，具体目标有三：

目标一：实现对核心制造设备的数据采集和云端 3D 数字孪生。

目标二：利用大数据分析实现核心制造设备的产能监控、故障诊断和耗材维护预测，生成相关统计报表。

目标三：构建生产设备资产管理应用服务，实现生产设备资产监控、管理和维护流程，对接企业 ERP 系统。

2、解决方案

本项目基于中国移动 OneNET 物联网开放平台使能，借助 OneNET 的后台大数据能力实现离线数据分析和实时数据分析，并由 OneNET 团队为客户定制相关企业应用服务。

1. 数据来源

(1) 物联网设备采集

为实现机床状态的实时监测，项目中为客户目标机床量身定做了相应的物联

网采集设备，通过接入机床电源和植入安装传感器的方式，可以准确实时采集目标机床的电流、电压、转速、机头位置、加工物尺寸、位置、环境温度、湿度和噪音等信息。

物联网采集设备通过本地网络或者蜂窝模组接入到中国移动 OneNET 物联网平台中，实时上报绑定机床的状态信息。

项目中，设备每 3 秒采集一次传感器数据，每 15 秒向 OneNET 平台上报一次数据记录。每台机床每次上报数据量约为 1KB。

（2）企业 ERP 系统

项目中机床的资产信息、财务信息和耗材记录，通过企业 ERP 系统接口获取，注入到相应的虚拟机床设备数据记录中，该虚拟机床设备与机床上的物联网采集设备绑定。

（3）车间报表

机床的生产记录、维护记录和故障记录等信息，客户原来通过车间纸质报表的方式记录。项目中通过车间报表采集录入相关数据，并注入到相应虚拟机床的数据记录中。

2. 技术方案

（1）系统架构

本项目软件系统分为企业应用服务、数据访问服务、数据分析服务、物联网应用使能服务和物联网接入服务等几大部分，具体如图 1 所示。



图1 核心机床设备资产监控系统

（2）企业应用服务

企业应用服务是项目中直接面向客户的应用服务，主要从机床设备资产管理和监控的角度为客户提供重要机床设备的产能监控、故障监控和实时监控服务，通过数据可视化和统计报表的方式为用户提供相关信息。

为了满足客户的需求，在企业应用服务中，系统实时采集机床数据并且建立了机床 3D 模型，为客户提供了基础的数字孪生功能，方便客户以远程方式实现机床车间的巡检。

另外企业应用服务还对接了客户企业 ERP 系统，根据产能监控信息、耗材预测信息和故障监控信息，实现相关的资产管理和维护管理流程。

（3）物联网服务

物联网接入服务基于 OneNET 物联网开放平台实现。机床状态采集设备采用 EDP 协议接入 OneNET 平台，根据配置的元数据模版，将时间戳，电流、电

压、转速、机头位置、加工物尺寸、位置、环境温度、湿度和噪音等信息以时间序列数据流的方式上报。

物联网应用使能服务同样基于 OneNET 平台使能。根据条件告警配置，平台提供位置围栏、环境温度、环境湿度和噪音的阈值告警。根据订阅配置，平台将收到的数据实时分发到实时数据分析服务进行流计算分析。

(4) 数据服务

离线数据分析服务基于 Hadoop 和 Spark 实现，通过 ETL 的方式导入 OneNET 采集的机床状态历史数据、车间生产记录数据、车间故障记录数据和企业 ERP 系统的资产信息和耗材数据进行数据挖掘，获取生产状态判决模型，故障诊断模型和耗材维护预测模型。

实时数据分析服务基于 OneNET 实时计算引擎实现，以流计算的方式接收并处理 OneNET 采集的机床实时数据，根据计算模型进行分析并将分析结果写入实时数据库中，用于后续用户实时交互式查询和统计。

数据访问服务用于为客户提供安全的标准化的数据访问接口。客户系统可以通过数据访问服务查询实时分析或者离线分析后的结果数据，或者导入要进行离线分析的数据集和元数据定义。

3. 项目其他亮点

本项目的亮点功能主要有：

(1) 3D 虚拟机床巡检，利用了 3D 建模技术并结合实时状态实现了机床 3D 动态显示，方便管理人员在线巡检。

(2) 基于流计算的实时生产状态判决、故障诊断和耗材维护预测，提升了资产利用效率，优化了耗材库存。

3、实施效果与推广意义

客户企业在应用本项目解决方案后，有效实现了提高生产设备资产利用，达到了降本增效的目标。

依据项目提供的产能报表，客户企业在小范围区域内试进行了资产优化，有效降低了电雕机产能闲置，减少了 5%~8%工厂的电雕机浪费，提升了现有电雕机 10%~15%的利用效率，并削减了当年的电雕机采购计划。

依据项目提供的资产维护功能，客户在试点工厂发现了耗材的异常使用情况，降低了 20%左右的耗材库存并减少了耗材虚报现象。

保守估计，项目解决方案全面部署后，每年将为客户增加约 3000 万利润。

4、案例亮点

1. 项目先进性及创新点

(1) 数字化孪生

项目初步实现了电雕机设备的云端数字化孪生，支持多类资产信息和状态信息的查阅，并提供 3D 虚拟视图，如图 2 所示，可以直观快速的查看到电雕机实时的生产状态数据以及工作情况。



图2 电雕机 3D 虚拟视图

(2) 用大数据分析实现设备隐态监测

项目借助大数据分析和机器学习技术实现了对电雕机生产状态、耗材消耗状态和故障状态等传感器无法直接感知状态的监测。

六、 联想工业大数据平台 LEAP

1、 案例背景与业务痛点

全球工业正面临深刻的变革，一方面工业企业日益关注小批量个性化生产，争夺快速增长的“用户定义制造”市场蓝海；另一方面，新技术的飞速发展以及各个领域的快速渗透，使得工业企业的传统模式变革以及新型业务模式创新成为可能。为了实现工业转型升级，工业企业需要推动自身业务系统和流程的全面升级，在这个过程中需要解决如下挑战：

(1) 企业内多个异构系统间的数据无法有效整合,直接导致企业采购、生产、物流、销售等环节割裂，效率降低；

(2) 企业无法对生产设备进行实时数据采集和统一灵活控制，导致企业难于实现生产工艺流程的最优化，以达成大规模个性化定制；

(3) 随着海量新旧数据的不断积累沉淀，企业需要可靠的低成本方案提高数据存储和计算能力，实现对海量数据的高效管理；

(4) 企业智能化分析门槛高，难于整合分散在业务中碎片化领域知识，实现跨业务和跨领域的业务流程再造；

(5) 在实现数据价值变现的同时，企业也必须构建基于硬件的大数据安全防御体系，保障数据资产和核心工业流程的安全。

只有通过构建企业级工业大数据平台，构建产业生态，实现多个工业软件的云化协同，才能为网络众包、协同设计、大规模个性化定制、精准供应链管理、全生命周期管理、电子商务等新模式下的企业生产经营带来价值链体系重塑。

2、 解决方案

联想工业大数据解决方案包含了大数据智能平台 LEAP AI、大数据计算平台 LEAP HD、物联网采集及边缘计算 LEAP Edge Server、数据集成平台 LEAP DataHub、数据治理平台 LEAP DataGov 和可信计算引擎 LEAP Trusted 等产品线，

包含了数据整合、计算引擎、数据分析算法和模型、数据治理、数据安全保护及行业解决方案等各个层次的服务。实施概况如下。

1. 数据来源

(1) 关键工业设备数据

该数据包括工业传感器、数控/模拟机床，工业机器人，产线检查设备，现场监控设备等工业终端数据。

(2) 工业系统数据

该数据包括 ERP、CRM、MES、SCM、PLM 等系统数据。

(3) 外部数据

该数据包括网络爬虫抓取的互联网数据、设备及网络获取的用户和社交数据、权威机构发布的产业数据等。

2. 技术方案

(1) LEAP 大数据平台整体架构

联想 LEAP 工业大数据方案包括三个功能层次：

其一，LEAP 平台提供不同技术手段保证了企业内外部数据的高效联通，其完善的数据集成工具支持对多源异构数据的高效集成与处理，工业物联网采集及边缘计算能力能够实时采集企业设备数据及生产数据；

其二，基于 LEAP 产品家族，联想构建了企业统一数据湖方案，可以帮助制造企业高效融合 OT，IT 以及 DT 数据，打通制造企业内部的关键设备与工业系统中的数据孤岛，以私有云、公有云或混合云的方式实现企业内部的数据互通和与外部关联企业间的知识共享；

其三，根据不同制造业细分领域客户的应用需求，LEAP 提供了丰富的、可集成的行业应用集合，通过 LEAP 产品家族的行业算法库快速构建分析模型，提供制造流程中关键场景业务优化能力。

其技术架构如下图所示。

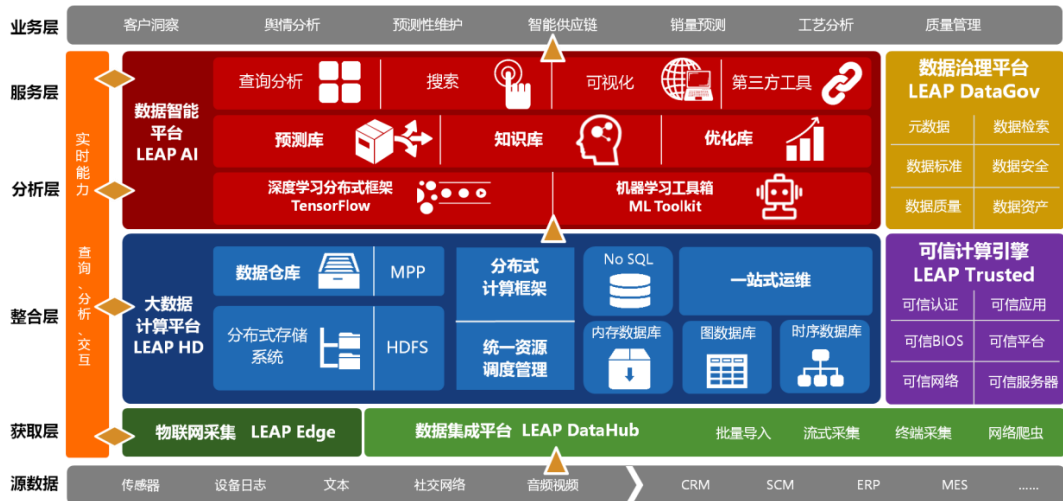


图3 联想工业大数据方案技术架构

(2) 数据集成平台 LEAP DataHub

联想数据集成平台 LEAP DataHub 支持对多源异构数据的高效集成与处理。它支持批量、流式、网络爬取等多种数据采集方式，支持各类数据的 ETL(抽取、转换、加载)过程，支持多种任务调度方式，以满足不同的数据处理需求，并且能够根据企业的需求快速扩展。LEAP DataHub 覆盖 50 余种主流数据库/数据接口，能满足企业在复杂业务场景下的各类数据整合要求。同时，它提供全图形化的数据处理工具，通过拖拽方式设计各类 ETL 过程，简便易用。

(3) 工业物联网集成平台 LEAP EDGE

联想工业物联网集成平台 LEAP EDGE 帮助用户从物联网大数据分析中获取最大业务价值。它提供强大的多源异构的海量数据采集与整合能力，支持多种物联网设备和工控系统的采集方式，能够根据企业的需求方便地快速扩展，支持海量、多样的物联网数据的接入、集成与分发。同时 LEAP EDGE 具备实时的物联网大数据分析能力，能够通过实时采集、实时处理，相关的分析规则和分析算法，可以进行实时分析、实时预警。

(4) 大数据计算平台 LEAP HD

联想大数据计算平台 LEAP HD 是整个大数据存储处理和分析的核心基础平台。它基于 Hadoop/Spark 生态系统，引入了多种核心功能和组件，对复杂开源技术进行高度集成和性能优化。在分布式存储系统的基础上，建立了统一资源调

度管理系统，深度优化大规模批处理、交互式查询计算、流式计算等多种计算引擎。LEAP HD 整体性能超群，具有海量数据实时处理能力，支持物联网实时业务分析，具有使用简便、运行高效、易于扩展、安全可靠等特点。

（5）数据智能平台 LEAP AI

联想数据智能分析平台 LEAP AI 提供深度学习分布式框架、机器学习工具箱、预测库、优化库、知识库等建模工具，具备特征工程、数据建模以及机器学习算法学习库的功能，可以辅助用户发掘隐藏在数据背后的巨大商业价值，加快从数据到业务的价值实现。系统支持 50 多种分布式统计算法和机器学习算法，不仅提供传统数据挖掘算法，还提供了自然语言处理、文本分析、水军识别、信息传播等原创前沿机器学习组件。除此之外，联想对算法精度进行深度优化，优化后的性能比开源算法库提速 3—10 倍。

（6）数据资产管理 LEAP DataGov

数据资产管理 LEAP DataGov 将数据对象作为一种全新的资产形态，围绕数据资产本身建立一个可靠可信的管理机制，提供数据标准管理、数据资产管理、元数据管理、数据质量管理、数据安全等功能，为数据管理人员、运维人员、业务人员和应用开发者提供全方位服务与支撑。

（7）可信计算引擎 LEAP Trusted

联想可信计算引擎 LEAP Trusted 是联想基于自身多年安全防护实践经验和对企业级复杂安全业务需求的充分了解所打造的安全可信产品。基础硬件平台采用基于 TPM/TCM 可信技术的硬件 Server，从硬件到 BIOS，再到 OS，再到大数据平台，进行逐级可信验证，确保整个平台的可信安全。同时提供可信接口，可以实现对第三方应用的可信验证。对数据进行整个生命周期的安全管理，包括数据的安全采集、数据的安全存储、数据的分析挖掘、数据资产管理以及运维服务等。实现全体系监控，提供用户日志、行为积累以及大数据平台的审计。

3. 项目其他亮点

联想依托自身软硬件一体化的优化能力，全球化的业务能力，打造了开放的、

可信的企业级一站式大数据平台，为解决企业大数据问题，提供了一站式的解决方案。通过联想大数据平台，可以轻松完成异构数据、分散数据的整合，实现企业内部分散数据和外部数据的融合，进行供应链、客户经营、产品设计、质量等方面的优化，快速发掘隐藏在数据背后的巨大商业价值。

3、 实施效果与推广意义

1. 联想设备销售激活分析

通过该大数据分析平台，联想集团完成了近 5 年生产、销售、物流和设备激活等数据存储和分析处理，实现了分地域、国家的设备销量的统计分析，生产和销售部门根据设备的区域销售情况及时合理调整生产和销售策略等重要功能。联想集团仅通过对印度地区的生产、物流和销售渠道的调整，便促进设备销量提升 18%，节省生产物流费用近千万美元。

2. 联想设备全生命周期管理

利用大数据，联想首次实现了 1~2 天完成一次产品质量迭代，远远优于基于传统方法一至几个月的迭代周期。每年节省六百万美金的设备维修费用，新版本发布速度提升 6~10 倍，产品投诉率下降 63.6%。

在采用大数据解决方案之前，提升产品的质量最大的问题在于数据来源少，例如一个典型的质量优化流程，从用户发现缺陷，到最终技术人员解决缺陷并发布到用户设备上，往往需要一个甚至几个月，大大降低了用户对产品的满意度。而通过利用大数据技术，联想可以通过在全量移动设备上的数据跟踪，实时/非实时获得产品软硬件数据，捕获产品各类异常问题。

3. 某钢铁企业需求预测分析

通过大数据分析系统，整合了内外部各类相关数据，借助机器学习和知识图谱来发现和探索，发掘出数据和业务之间的联系。相对于过去采用首席专家凭经验预估市场需求量，机器智能预测的方案大幅度提升了预测准确率和客观程度。

针对主要 6 个区域，利用历史数据为评测指标，针对给定的某一个历史时间

点，预测以它为起点一个月后的需求量相对误差 $<15\%$ ，6个月后的相对误差 $<30\%$ ；针对主要6个大客户，利用历史数据为评测指标，针对给定的某一个历史时间点，预测以它为起点一个月后的产量相对误差 $<20\%$ ，6个月后的相对误差 $<35\%$ 。

4、 案例亮点

联想工业大数据平台依托于联想软硬件一体化的优化能力，全球化的业务能力，通过深入优化开源和硬件创新，打造了一个开放的、可信的全球大数据和云计算的基础业务平台，并帮助中国骨干企业实现全面的大数据业务能力构建，实现了企业的管理和创新能力提升，具有很高的应用价值和行业示范效应。联想大数据平台涵盖大数据领域的多个核心技术，并构建了面向骨干企业的供应链优化、客户经营、产品优化、质量控制等诸多方面的全面大数据和云计算创新行业解决方案能力，有力的支持了产业转型升级和企业间数据合作。联想工业大数据解决方案通过在生产设备实时分析、多源异构管理系统整合、可信数据安全、领域知识场景化、自动流程管理等方面的自主技术创新，构建了覆盖全球的大规模工业大数据实时分析集群，并成为国内制造领域的最大工业大数据集群实践。

七、 格力电器工业大数据平台

1、 案例背景与业务痛点

珠海格力电器股份有限公司是一家多元化的全球型工业集团，主营家用空调、中央空调、智能装备、生活电器、空气能热水器、手机、冰箱等产品，2017年格力电器实现营业总收入1500.20亿元。

格力集团在多年的发展中积累了大量的生产经营数据及丰富的数据分析经验，格力集团通过与航天云网合作，建设具有格力集团特色的工业大数据平台，通过大数据分析企业运营状况，支撑集格力家用机、商用机等主流产品的设计、

生产、物流、销售、服务为一体的智慧企业的运行，完成格力电器智能化转型。

格力工业大数据平台通过生产经营及智能产品数据的集成、处理、分析、挖掘等，实现设备故障诊断、故障预测、产品统计、实时查询、营销支持、智能搜索等功能，将互联网、大数据、人工智能技术与家电制造业融合，实现企业经营精准化、生产现场透明化、产品服务远程化等目标，实现企业数据驱动的持续优化。

格力工业大数据平台的核心功能包括：

- 1、实时数据接入大数据平台，离线、在线数据的接入；
- 2、智能业务系统数据互通和智能化分析；
- 3、基于机器学习的智能大数据分析；
- 4、工业大数据语音检索；
- 5、工业大数据平台的管理与运维。

2、 解决方案

格力工业大数据平台以在线实时同步的方式接入了集团内部的 ERP、MES 等业务系统数据，同时以在线及离线的形式接入了上百万台设备的测试、调试、运行数据，数据总量在千亿级以上，数据接入速度达到 8.66 万条/秒、5.65G/分钟，数据处理速度达到 9.58 万条/秒、6.25G/分钟；智能平台通过后台智能运算和机器学习，实现了工业生产各个环节的自动预测协调和资源的智能配置，构建了数据驱动的现代化智慧企业。平台的功能架构如图所示：

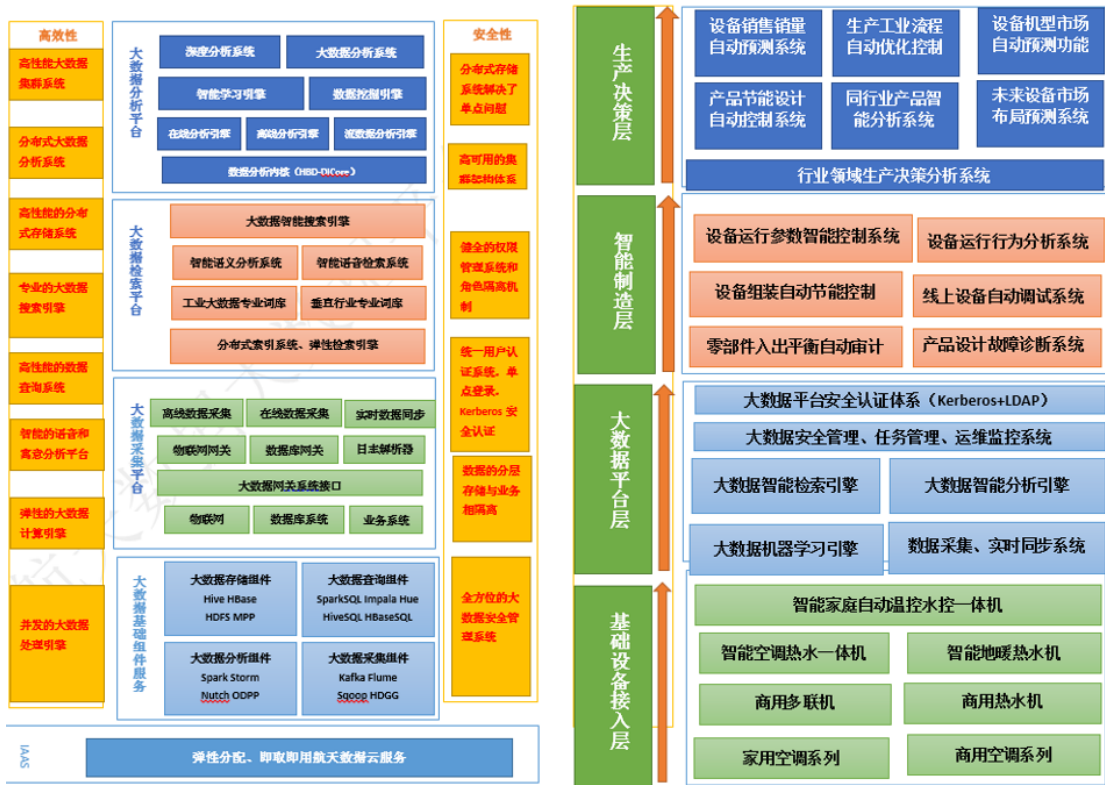


图1 格力工业大数据平台功能架构

数据采集平台是整个大数据平台的数据源，负责对接所有的外部数据。在实际环境中，外部接入的数据源种类繁多，包括来自生产设备采集的非结构化数据，来自日志系统等的外部文件数据，各信息系统原有的关系型数据库接入和实时的流式数据。为了平台更好的兼容性和扩展性，数据采集平台采用多种接入策略。实时数据可通过 kafka 作为消息服务接入，非结构化的互联网数据、文件数据和传统的关系型数据库数据可通过 Apache Flume 接入，并对原始数据做一定的初级数据清理加工处理。

大数据分析管理平台是整个大数据平台的核心，承担着数据存储、数据分析的核心功能。主要核心技术基于 Hadoop 生态圈，在基础的分布式文件系统之上 (HDFS)，主要由存储系统和计算系统构成。针对数据特点，存储系统由 HDFS 和面向列存储的 Hbase 构成。计算系统利用 Hadoop YARN 做统一的资源管理，基于内存的 Spark 和传统的 MapReduce 结合，充分利用两大计算框架的优势和资源。Mahout 和 Spark MLlib 提供了丰富的数据挖掘、机器学习服务，Spark SQL 有着强大的关系型数据处理能力，Spark GraphX 可完美应对大规模图计算的需

求，Spark Streaming + Storm 的解决方案可以满足实时计算需求。

在大数据分析管理平台之上，可通过统一的数据总线与上层应用系统和其他服务对接。其他服务对整个大平台的支撑起重要作用，包括业务数据的存储管理、分布式的缓存系统、运维平台的监控告警等等。通过与大数据分析管理平台的交互，可以灵活构建丰富的上层数据应用，比如数据质量的监控、关联分析、语义分析、自动的模型建立与优化、知识库指标库的构建等。

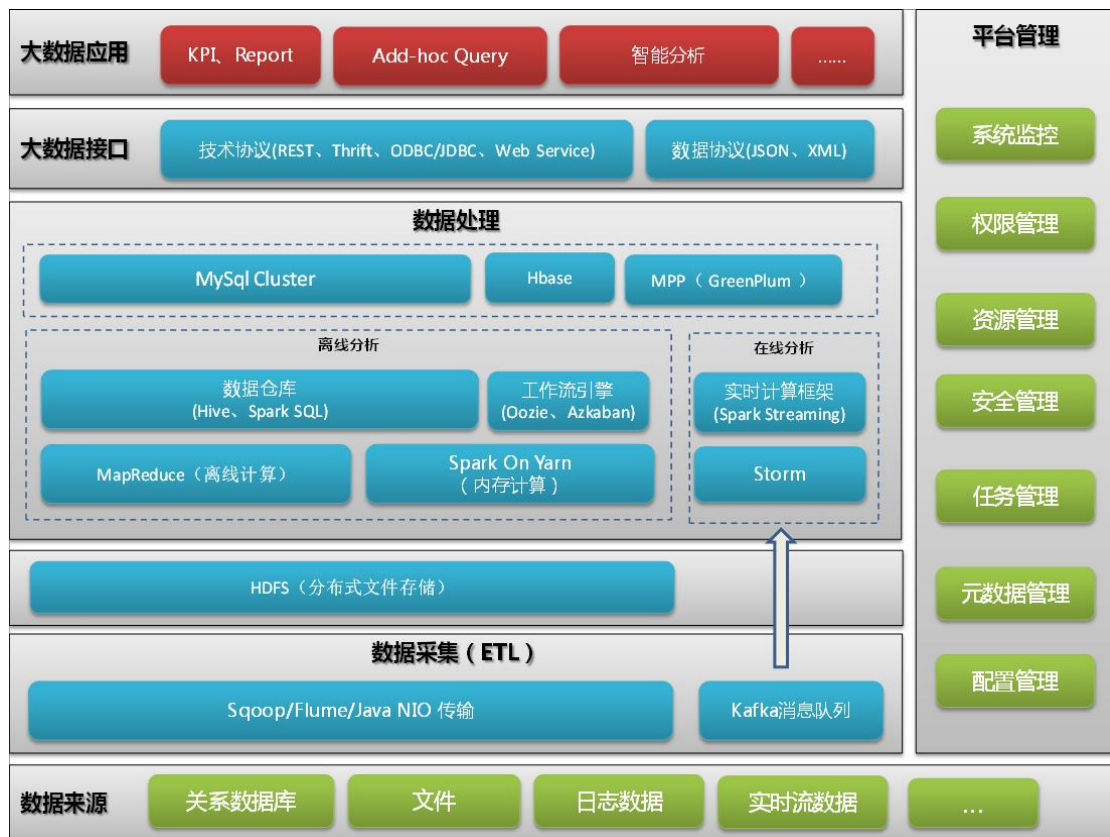


图2 格力工业大数据平台技术架构

1、大数据采集技术

大数据采集系统包括非结构化数据及结构化数据的采集，非结构化数据采用 FlumeNG 采集, Oracle/SQLServer/MySQL 关系数据库中的结构化数据采用 Sqoop 收集。此外还有以下实时流数据采用分布式消息队列采集。

(1)Flume 收集非结构化数据

Flume 是高可用、高可靠、分布式的海量数据采集、聚合和传输系统，Flume

支持定制各类数据发送方，用于收集数据。同时，Flume 提供对数据进行简单处理，并写到各种数据接受方的能力。非结构化离线数据写入 HDFS 流程，采集层生成压缩文件到磁盘目录，采用 Flume spooling source 监听文件目录，每 500ms 扫描一次文件目录，有新产生的文件时，立刻读取文件入库。

(2) Sqoop 收集结构化数据

Sqoop 用于在 Hadoop 与传统的数据库间进行数据的传递，将关系型数据库中数据导进到 Hadoop 的 HDFS 中，必要时也用于将 HDFS 的数据导进到关系型数据库中。Sqoop 使用元数据模型来判断数据类型并在数据从数据源转移到 Hadoop 时确保类型安全的数据处理，Sqoop 通过分割数据集并创建 Hadoop 任务来处理每个区块。

(3) Kafka 收集实时数据

Kafka 作为一种高吞吐量的分布式发布订阅消息系统，在格力工业大数据平台中用于处理消费者规模的所有动作流数据。这些数据通常是由于吞吐量的要求而通过处理日志和日志聚合来解决。对于像 Hadoop 的一样的日志数据和离线分析系统，但又要求实时处理的限制，这是一个可行的解决方案。Kafka 通过 Hadoop 的并行加载机制来统一线上和离线的消息处理，通过集群机来提供实时的消费。在实时数据处理上，通过采集层数据通过消息队列组件 Kafka 接入到 Spark Streaming 里，Spark Streaming 实时处理后把结果存到 Hbase 或 Mysql 等关系数据库中供用户查询。

2、大数据分析技术

大数据分析系统以机器学习、数据挖掘等作为核心技术，构建于大数据管理系统和云计算平台之上。其中，大数据管理系统提供数据的存储与查询功能，云计算平台提供分布式并行计算服务。通过分布式计算与统计分析服务器访问大数据系统，实现 KPI 与报表统计分析服务。

格力工业大数据平台采用 Mahout、Spark MLlib 构建在 Hadoop 上对大数据进行挖掘处理，实现故障预测、配件库存优化、报表分析、日志检索等深度数据挖掘。在项目中，我们通过 Hadoop 系统 MapReduce 实现数据挖掘引擎的并行计

算。在使用 MapReduce 模型进行大规模数据处理时,重点是编写 Map 和 Reduce 函数,其它并行计算中的复杂问题诸如分布式文件系统、工作调度、容错、机器间通信等由 MapReduce 系统处理,提升了系统研发的效率。

3、大数据管理技术

大数据管理系统具有高可靠的架构设计,完全分布式的、多副本机制的、对等的、不共享系统,没有单点故障或瓶颈。这使得系统线性增长,每新加一个节点可以同时增加系统的性能和存储容量,具备好的扩展性能。

系统具备以下特点:扁平化设计,弹性扩展,系统采取扁平化设计,节点之间完全对等,都可以对外提供服务。柔性多引擎技术,对于不同的应用需求可以使用不同的引擎来对外提供服务,支持异构数据,结构化,半结构化,非结构化数据的统一检索。高效分区索引机制,可根据应用的查询特点,将数据自动分区索引,适应海量数据的集中索引和快速索引的应用需求,分区索引还可以减少检索时的索引匹配范围,缩短检索响应时间。混合索引方式,提供按词索引、按字索引、字词混合索引方式,满足不同应用场景对查全和查准的不同需求。异步检索,支持异步检索模式,适应大并发的应用场景要求,避免了同步检索模式时消耗太多线程资源的问题。

3、 实施效果与推广意义

格力集团董事长兼总裁董明珠在某次大型讲座上表示,“我们用互联网、大数据做空调,格力空调运行得怎么样,我们在珠海就可以监控到相关数据。”截至 2017 年 12 月,格力多联机空调每天返回数据处理中心的机器运行数据超过 1.5 亿条记录,目前采集数据超过 100T,格力通过故障数据分析指导工程安装,提高工程安装质量,欠氟和漏氟故障率下降 22.5%,电子膨胀阀故障率下降 21.3%。总体来看,格力工业大数据平台实施效果及意义主要有以下几点:

1、故障诊断:

大数据平台处理、分析商用空调运行时采集回传的工况数据,及时、准确定位运行故障并给出大致原因,为维护部门维修空调设备和系统提供信息支撑,降

低故障定位的时间、范围和工作量，缩短停机时长并提高客户满意度。

2、故障预测：

大数据平台利用数据挖掘、机器学习技术，通过学习已存在的设备故障数据、信息，尤其是发生故障的前兆数据，归纳故障发生的特点、规律知识，并利用流计算相关技术及时发现潜在的故障及风险，及时预警，减少停机的次数及停机时长。

3、实时查询：

大数据平台利用 New SQL 存储技术存放设备数据，提供特定编号设备、一段时间内工况信息的实时查询功能，用于业务人员分析、判断特定设备在查询时间范围内的工作状态，总结、发现业务规律。

4、营销支持：

大数据平台利用集成的内外部产品及销售数据进行统计分析，挖掘其中的重复购买、交叉购买等特定行为的规律，分析产品销售情况，客户行为信息，特定产品的销售规律、特点及变化趋势，指导业务部门进行产品的个性化设计及精准推广营销。

4、 案例亮点

格力工业大数据平台是航天云网工业大数据技术实践与应用典型案例之一，平台实现了设备数据的自动采集、实时同步、智能清洗、自动聚类、在线分析、数据挖掘、深度学习、分析展示、智能语音语义检索、用户行为分析、工业生产决策等功能，该平台的亮点可以概括为平台交互、设备运维、企业运行的“三个智能化”：

1、平台交互的智能化

采用 HTK 和 JuliusJS 组件将语音转换成文本，采用 mmseg4J 技术对语言文本进行分词，通过应用交互式探查分析，根据搜索内容，推测查询目的，根据历史机器学习数据，即时构造查询，系统还可以根据查询内容进一步实现查询结果

的过滤、分析和保存等操作，实现企业大数据平台交互的智能化。

2、设备运维的智能化

大数据平台通过对商用空调历史运行数据的挖掘与学习，归纳故障的特性与规律，建立设备故障信息模型。在设备运行过程中，通过流计算等技术对设备的潜在风险进行实时预警。设备运行过程产生的故障模型数据，如停机模式、相关规律等，定期形成报告推送至产品设计部门、生产制造部门业务系统中，对产品设计优化、生产工艺优化提供数据支撑的指导意见。

3、企业运行的智能化

格力工业大数据平台高效整合企业信息化数据、设备物联数据及外部关联数据等海量异质异构数据源，通过全面数据汇聚分析，在生产现场有效保障设备的健康运行，在企业经营管理领域提供商业决策支撑，在产品服务端实现设备的实时监控和预测维护。在企业全方位数据感知的基础上提供合理优化建议，推动企业经营管理的智能化转型。

八、合力叉车工业互联网平台

1、案例背景与业务痛点

叉车是工业搬运车辆，国际标准化组织 ISO/TC110 称为工业车辆，广泛应用于车站、港口、机场、工厂及仓库等各个国民经济部门，是机械化装卸、堆垛和短距离运输的高效设备。从叉车应用范围来看，其应用领域很广，几乎包括国民经济方方面面，因此，整个行业景气度与宏观经济息息相关。

安徽合力股份有限公司是中国工业车辆的领军企业，已连续 27 年排名行业第一；2011 年进入了福布斯中国 500 强；2016 年进入世界工业车辆行业七强。

合力自 1992 年开始了企业信息化的征程，20 多年来积累了大量的工程与运营数据。2011 年启动了以信息化引领管理变革、以大数据应用推动运营转型的第

五次管理革命，陆续在总部、各网点和分子公司全面上线运营管理与数据分析系统；2014年，启动实施了以整车宝、维修宝、报工宝、收发宝等为代表的十八项具有自主知识产权的工业互联网APP，囊括了销售、生产制造、采购供应、售后维修等所有环节，实现了传统企业生产运营模式工业互联网升级；2016年以传感、通信、机器自感知和数据采集为核心的叉车车队管理系统，有效为客户在驾驶叉车时提供实时分析、维修预告和精准路线服务；2017年合力自主研发的无人驾驶叉车，成功实现自动出入库作业，集成调度系统、货物识别装置等软硬件，成功实现了无人化、智能化仓储物流搬运作业。

安徽合力以改革促转型、创新真升级，应用移动互联网、云计算、大数据等技术，抢占工业车辆行业“制空权”，服务于工业车辆行业的市场营销、研发设计、生产制造、物流搬运、售后服务、车辆租赁和金融租赁等业务，通过在拥有自主知识产权的工业互联网大数据平台，为客户提供全面的物流搬运解决方案。

1. 项目背景及业务痛点

从行业的宏观背景上看，工业车辆行业的企业现在面临的竞争环境已与过去完全不同，多品种、中小批量、多配置代替了过去的标准化产品，多样化市场从统一的市场中迅速成长，产品生命周期和开发周期日益缩短。

过去，企业通常要么追求低成本，要么追求品种多样化，而如今他们则越来越明确地认识到，必须采取能够同时实现效益生产和小批量定制化生产的企业策略。为了不再大规模地生产标准化产品或者高成本地生产多品种的产品，企业会发现，可以通过“模块化设计，标准化生产”的生产方法，同时达到产品的低成本和品种多样化的目的。简言之，多品种、中小批量、多配置模式是指对标准化的产品和服务进行个别的规模化生产。它将“标准化”与“定制化”这两个表面上矛盾的概念组合在一起，其背后隐藏着极其深刻的逻辑关系，并将给当前众多工业车辆行业企业的经营运作与管理体系带来巨大的冲击与改变。

在以客户、竞争和变化为特征的新经济时代，企业要想在市场竞争中获得竞争优势，以确保企业的生存和长期发展，都必须回答这样一个问题：如何满足并快速响应客户的个性化需求？为了找到正确的答案，企业需要正视以下这些挑

战，并在新的“模块化设计，智能化生产”模式中得到解决。

- 1) 需求的分化，统一的大市场已日益多元化。统一的市场变成很小的细分市场，市场正在向这样的顾客转移——他们需要高质量并且更加接近他们个性化需求的品。
- 2) 个性化产品的需求是不稳定的。过去对标准产品的大量需求已分割成对相似产品不同“风味”的需求。
- 3) 按以前的生产方式已不能保持原有的利润，因此，追随具有多样化需求的细分市场并尽量满足这些细分市场中不断变化的需求，似乎更为可取。使用以前的生产方式来定制面向细分市场的产品（通常通过服务）也可以做到这一点，但成本太高。而增加产品多样化最终还必须通过生产来实现。
- 4) 在生产中建立丰富的多样化，不能通过专用的规模化生产技术来实现。实现产品的多样化，需要制造过程的灵活性和适应性，这与规模化生产方式是矛盾的。
- 5) 生产模式必须改变。现代生产模式是由市场和客户驱动的，必须以很短的生产周期生产大量不同品种、高质量的产品。这需要有柔性能力的制造装备和较高综合技能的工人。
- 6) 由于新产品极大程度地满足了客户的需求，所以一般可以取得一些额外的收益。这样得到的利润差弥补了由于产量低带来的低效率。随着在规模化定制过程中经验的积累，会经常发现具有多品种的产品能够以与标准产品相同或更低的成本生产出来。
- 7) 由于细分市场越来越小且不断变化，只有以更快的速度生产出更多品种的产品才能不断取得成功。产品技术变化的速度日益加快，因此必须以同样的速度缩短产品开发周期。
- 8) 随着产品开发周期的缩短，产品生命周期也在缩短。为了最大程度地满足客户的需求，产品和技术应不断地改进和更新。

2. 项目简介

合力工业互联网平台是通过数据+模式的方式为用户提供服务，囊括了销售、生产制造、采购供应、售后维修等所有环节，实现了传统企业生产运营模式工业互联网升级；应用移动互联网、云计算、大数据等技术，服务于工业车辆行业的市场营销、研发设计、生产制造、物流搬运、售后服务、车辆租赁和金融租赁等业务；结合企业车联网数据感知体系建设、车联网云平台建设以及 AGV 无人驾驶体系建设。

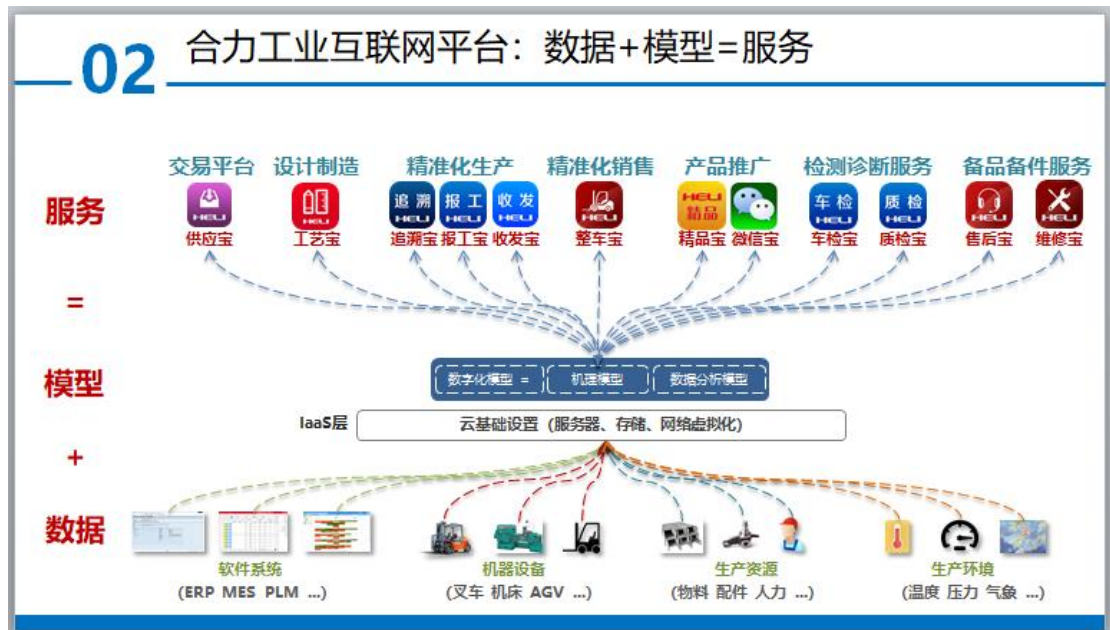


图1 合力工业互联网平台服务模式

通过叉车车载终端设备，如定位设备、重量传感器、红外传感器、胎压监测仪等，结合无线通讯与企业网/互联网，将基于 CPS 系统的车辆运行情况进行采集、处理、分析、统计，从而实现叉车的监控、调度、管理等功能的跨行业跨领域综合工业互联网平台。

该平台可实现远程采集数据协助合力对叉车的研发提供数据，以及对车队系统进行升级；帮助企业更好的使用合力叉车，提高安全性、效率及经济性。同时优化整个供应链体系，促进关键体系的综合集成、协同与创新，业务的紧密协同，为合力及相关方提供全面的移动互联网与工业融合的工业车辆生态圈解决方案。

3. 项目目标

合力工业车辆互联网平台，为工业车辆行业实现五大目标：

第一，从面向信息化管理的应用平台转为智能制造的支撑平台。通过各种各样以 SaaS 软件和工业 APP 形式呈现出来的服务，能够提供从单体叉车、到生产线、到产业链、再到产业生态系统的系统级优化，实现从局部优化到全局优化。

第二，从以往的单一功能平台变为创新驱动引擎。在工业里很多技术、知识、经验、方法创新需要从 0 开始，知识复用水平较低。而构建一个工业互联网平台，将大量的工业技术原理、行业知识、基础工艺、模型工具、业务流程以及老专家脑子里面的几十年的经验进行规则化、软件化、模块化，以数字化模型的形式沉淀在这个平台上。沉淀完之后就不需要再做重复性工作，可以直接调用、复用、传播，重构我们的工业创新体系，大幅度降低创新成本和风险，提高研发、生产和服务效率。

第三，从内部封闭平台需要转为开放的生态平台。通过构建精准、实时、高效的数据采集互联体系，建立面向工业大数据存储、集成、访问、分析、管理的开发环境，实现工业技术、经验、知识的模型化、标准化、软件化、复用化，不断优化研发设计、生产制造、运营管理等资源配置效率，形成资源富集、多方参与、合作共赢、协同演进的制造业新生态。

第四，从由服务合力转为服务生态链客户。合力发展需要构建生态链，可持续性的良性收益更是建立在以客户为本的各个环节配合之上，高质量、强依赖度的服务，是企业生态链中最重要的一环。合力需要整合“平台提供商+应用开发者+海量用户”生态资源，抢占工业大数据入口主导权、培育海量开发者、提升用户粘性，构建基于平台的制造业生态链，不断巩固和强化制造业竞争优势，实现从成本中心向效益中心的转变。

第五，实现重点领域关键技术自主知识产权化。从当前现状来看，我国 AGV 发展历程较短，国外厂商的 AGV 技术已日趋成熟。合力的 AGV 导航控制模块由于采用了国外厂商的解决方案，为了降低产品成本，实现国产替代，提升企业核心竞争力，需要在核心技术上加快研发和消化吸收，逐步摆脱对国外厂商的依赖，全面实现国产化。

2、解决方案

平台建立基础大数据平台及系统优化，搭建大数据存储相关计算平台，大数据分布式系统，通过大量的工业 APP 应用和车联网实现数据采集和敏捷可视化，通过工业车辆的无人驾驶实现智能制造和智能物流方案。

1. 数据来源

(1) 用户需求数据采集

通过移动设备与客户进行产品推荐、视频展示及需求反馈，根据客户浏览行为，将客户进行记录分类，为后续精准销售提供数据支持。

(2) 销售订单数据

销售人员通过移动设备进行订单录入，含车辆基础配置信息、用户相关信息等，上传至云平台，通过接口转至生产系统进行排产，后期可以在移动设备中查看生产及运输进度，与生产系统无缝集成。

(3) 物料转移数据

生产过程中频繁出现物料的移动，对物料每一次移动进行记录上传至云平台，为精准库存管理提供基础数据支持。

(4) 加工过程辅助数据

进行零部件加工时，通过云平台提供的与设计系统的接口按需调用设计相关图纸及数据或变更信息，有助于提高加工精确度。

(5) 生产进度数据

提供移动端及客户端多种方式，支持一线生产员工实时进行生产过程汇报，协助管理部门有效进行生产力分配，提高产能。

(6) 生产质量检测数据

对已经下线的车辆按质量管控要求进行质量检测，将检测数据上传至云平台，形成质量检测数据库，辅助生产过程质量管控。

(7) 车辆关键零部件数据

车辆生产及组装过程中，对涉及到的后期维护或保养或确定为车辆的关键零部件的相关信息通过扫码枪或其他扫描设备进行采集上传至云平台，车辆生命周期全过程均可通过云平台接口进行调用。

(8) 车辆性能检测数据

车辆在出厂前需要经过一系列性能检测，根据研发及实际工况要求，通过实验台对相关数据进行采集分析整理，上传至云平台车检数据库完成基础数据采集。

(9) 售后维修数据

在用户现场通过语言、文字、图片实时进行维修数据上传，在云平台形成维修数据库，可以进行后续相关产品质量或其他维度分析，为产品后生命周期管理提供原始真实数据。

(10) 新产品试验数据

新产品批量上市前需要经过大量多方面实验验证，通过云平台记录新产品各方面试验数据，进行汇总分析，对新产品进行多层面考证，为批量稳定上市提供准确数据分析支持。

(11) 工业车辆运行的实时状态数据

内燃和电动叉车的实时工作状态，历史数据等信息，通过叉车上安装的传感器和 CAN 总线的数据流，对采集到的数据进行汇总并按照基于 TCP/IP 为基础的底层通讯协议，发送原始数据至服务器，解析后在平台展示。

(12) 工业车辆运行时的报警信息数据

采集安装在整车上的传感器数据，内燃整车采集油量、碰撞、超速和驾驶员离位报警信息；电动叉车，除电量、碰撞、车速和驾驶员离位信息外，还包括电池终端单独采集的电流、电压等信息，并根据相应算法，判断电池的充放电状态。

2. 技术方案

(1) 总体架构

平台通过建设物联网平台、业务层、数据层和应用层，来形成整个工业互联网的平台架构。物联网平台提供基础支撑和物联网通用能力，业务层提供物联网应用设备能力，数据层提供数据存储、分析和应用集成等能力，应用层提供工业应用的接入能力。

整体方案架构

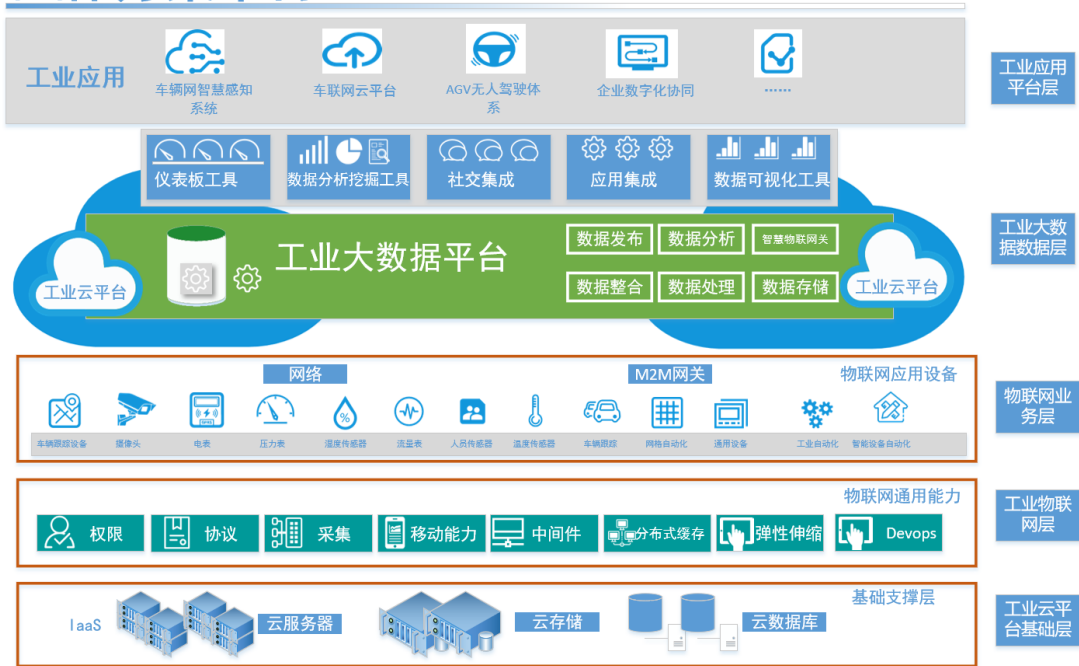


图2 合力叉车工业互联网平台整体方案架构

其中的工业大数据平台实现了工业大数据的采集、交换、清洗与集成。数据来源既包含来自传感器、SCADA、MES、ERP等内部系统的数据，也包含来自企业外部的数据。平台具有以下功能：

- 大数据清洗与存储**：首先针对采集数据进行数据清洗工作，对数据进行重新审查和校验，具体包括：一致性检查以及无效值和缺失值的处理。一致性检查是根据每个变量的合理取值范围和相互关系，检查数据是否合乎要求，发现超出正常范围、逻辑上不合理或者相互矛盾的数据并剔除；无效值和缺失值的处理方法有：估算，整例删除，变量删除和成对删除等。最终将脏数据转化为满足数据质量要求的数据。然后根据数据的具体类型格式存储至关系型数据库或者非结构化数据库中。数据存储

是数据以某种格式记录在计算机内部或外部存储介质上的过程、方式和结果。通过非分布式数据库，对文本数据及图片数据进行存储管理；通过对数据读写操作的分布式处理，可提升数据库整体数据写入和查询性能，满足数据快速检索要求；通过批处理计算集群，实现数据离线式的挖掘分析处理要求；通过流式计算集群，采用内存数据库技术，实现数据在线分析处理。

- b) 数据挖掘：通过指定的规则和算法，从大量的数据中自动搜索隐藏于其中的、有着特定关系性信息的过程。针对具体业务需求，采用大数据与机器学习技术，通过合适的数学模型，对采集数据进行深度挖掘计算。如建立叉车维保数据模型，利用历史及实时数据对叉车全生命周期进行预测性健康管理；研发叉车智能调度算法，搭建任务环境知识库，针对具体任务能够提供实时分析规划，合理安排等。
- c) 资源池管理：将多个异构资源池整合成一个更大的资源池，进行统一管理和资源分配。资源池管理包括对资源池的扩充和缩减，暂停使用某些资源池等操作。平台设计充分考虑数据集中管理的要求，对资源池进行了分类和分级设计，通过资源的当前使用情况和历史使用情况形成性能曲线，并预测剩余资源的使用时限，能够提供参考依据，保证了资源池设计可行性，并根据业务实际需求进行资源配置，且有一定的预留，保证资源池的合理性和可落地性。平台的一种可行技术架构如下：

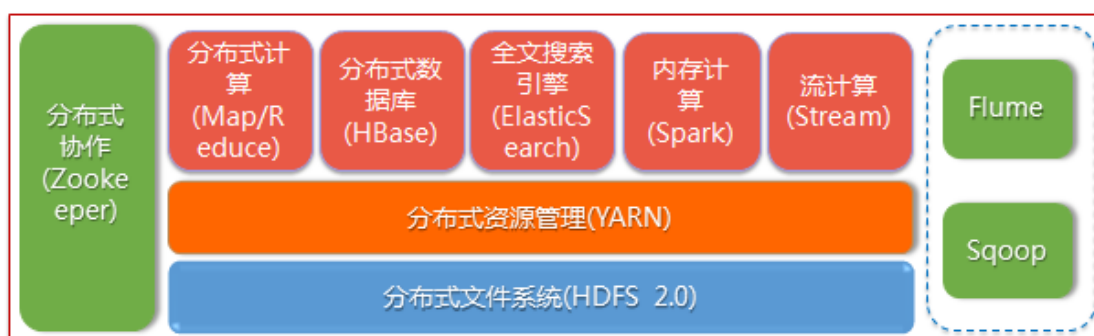


图3 平台使用的技术架构

(2) 主要组成

1) 工业物联网+车联网

工业车辆的传统模式已经受到大趋势的威胁，二十多年没有变化的销售，服务模式面临新的挑战。在新商业模式和资本的驱动下，创新颠覆模式正在孕育中。工业车辆企业要主动规划转型，以传统的叉车产品为中心的模式向以客户为中心的互联网模式转变，“互联网+”将更聚焦和关注工业车辆产业链，工业车辆上下游之间的跨界竞争。以叉车产品为载体，提供用户物料搬运服务。客户将来已经不再注重通过购买来获得产品的所有权，更倾向于获得产品的使用权，利用金融支持的工业车辆租赁和物料搬运服务将推向更高的层次。企业从卖产品，到卖服务，再到卖共用共享服务的模式。工业车辆企业将从产品市场份额的竞争转移到物料搬运服务市场份额的竞争，移动互联网让互联网共享工业车辆搬运服务成为可能。资本进入物流领域将力促更高效，更智能化工业车辆进入物流服务领域。电动汽车的分时租赁模式和专车的模式随着互联网+，也将进入工业车辆领域。

2) 工业云应用+工业 APP

通过构建基于工业 APP 的工业车辆制造服务生态圈，实现资源共享协同的生产组织方式，实现供应链协同；满足个性需求的制造，实现产品销售、设计和制造配置化模式；提升用户端设备，大幅度提高信息化覆盖率，打造智能绿色的生产运营；利用大数据进行行为分析，了解市场需求，实现精准营销；提供工业车辆融资租赁模式，发挥互联网金融参与度高、协作性好、中间成本低、操作便捷的优势；发挥工业 APP 对企业生产经营各环节的持续渗透与影响作用，利用工业 APP 实现企业内外全业务全流程互联互通、协作共享，提升生产效率和决策水平、降低成本、建立企业“互联网+”与“中国制造 2025”相融合的新优势。

3) 工业大数据体系

工业大数据的目的是为了改变以往工业价值链从生产端向消费端、上游向下游推动的模式，实现以客户价值为核心的定制化产品和服务，以及与之相适应的全产业链协同优化。为此，工业大数据应满足用户需求定义、工业智能制造、活动协同优化三方面的应用。工业大数据和互联网大数据的技术架构都具备数据环境、知识环境和应用环境三个层面。

从数据环境来看，工业大数据的数据环境，不但关注数据持久化，更关注数据采集的能力。工业大数据的数据采集依赖于物联网（IOT）的实现，IOT关注的采集的可靠性，实现数据接入的总线化，但工业大数据对IOT提出了更多的需要。

从知识环境角度，工业大数据需要对数据进行分析、处理，以获得相应的知识，用以支撑上层业务应用。工业大数据应用的模型相关性较强，工业大数据的知识环境的技术平台是CPS（信息物理系统 Cyber-Physical Systems），云计算是CPS的一个组成部分。CPS关注的是物理实体映射的逻辑实体的管理，提供逻辑实体的关系、协作，以对称的方式来演进，体现与物理实体的相关性，实现知识的挖掘。

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从机器设备实时采集数据，也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。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。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。

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。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。

3、实施效果与推广意义

1. 实施效果

- 1) 通过工业互联网平台为工业车辆用户提供智能调度等服务，实现从传统制造业向工业智能服务业的转变，实现全价值链的业务打通和数据共享，在本行业形成典型示范效应；
- 2) 通过工业互联网平台对车辆的监控预警，和在焊接、涂装等存在职业伤害的生产环境使用无人 AGV 作业，确保生产安全，降低职业伤害；
- 3) 通过本项目的实施，实现国内龙头制造企业、骨干设计院所、软件实施单位、核心设备厂家的集成，为中国智能制造自主化、集成化积累经验、培养人才、探索标准和规范；

2. 推广意义

- 1) 通过本项目的设备异常检测和预防性维保提醒功能，提高工业车辆可靠性，降低用户的使用成本；
- 2) 通过本项目的智能调度功能，合理分解组合任务、优化搬运路径，实现工业车辆的智能运维，提高用户的使用效率；
- 3) 自主无人工业车辆的研发应用，符合工业车辆发展趋势，对当前市场上的主流无人驾驶车形成较强的性能与成本优势，未来将提升企业经济效益和竞争力。

4、案例亮点

平台三大亮点：

1. 实现了工业车辆企业管理的移动可视化

平台基于工业 APP 的工业车辆制造服务生态圈，实现资源共享协同，通过生产类、销售类和管理类 APP，与企业资源计划管理系统（ERP）、产品全生命周期管理系统（PLM）、人力资源管理系统（HCM）、业务流程管理系统（BPM）、

制造执行系统（MES）、客户关系管理系统（CRM）、供应商关系管理系统(SRM)以及整车管理系统（VMS）等核心业务系统的高度集成，为企业市场营销、研发设计、生产制造、物流搬运、售后服务、车辆租赁和金融租赁等业务提供移动端应用。



图4 高度集成的可视化移动端应用

2. 实现了工业车辆产品的车联网化

平台基于车联网实现工业车辆车队资源共享、车辆状态实时掌控、叉车驾驶员信息、形成统计分析报表。通过底层硬件信息采集，GPRS 信息传输数据，人车卡绑定辅助智能报表分析，平台阈值设定下发至终端实现车辆控制，开关机状态实时监控与称重、行走状态合并运算实现车辆状态的实时监控，通过状态监控提取数据统计分析进而实现车队智能报表，平台各功能模块高度集成，形成工业车辆企业管理的车联网化，解决车队使用过程中，管理者对叉车自身状态不明、驾驶员工作情况的掌握不及时等问题，为叉车租赁市场、物流搬运等行业提供解决方案。



图5 工业车辆产品车联网化

3. 实现了工业车辆驾驶的无人化

平台实现合力自主研发的无人驾驶叉车，成功实现以 ERP、WMS、MES 等系统为上位系统，以上位系统调度任务的方式，通过无人叉车任务调度系统合理分配任务、优化行走路径，实现物料搬运的自动化、高效化、准确化搬运。配合多种自动化设备、传感器系统等实现物料的系统化管理，使得合力成为无人化、智能化的物流解决方案供应商。



图6 工业车辆驾驶无人化

九、东方国信大数据助力联合利华能源管理

1、案例背景与业务痛点

东方国信成立于 1997 年，是国家规划布局内的重点软件企业，公司成立 21 年来专注于大数据核心技术研发，2011 年在深圳创业板上市后，连续 6 年入选福布斯中国最具潜力上市企业榜单。凭借深厚的大数据技术积累和行业沉淀，东方国信已成为我国大数据龙头企业和工业互联网平台领先企业。

东方国信自主研发跨行业、跨领域的综合性工业互联网平台 Cloudiip，平台以助推中国制造 2025 为己任，以助力工业企业实现精益研发、智能生产、高效管理、精准服务为目标，已发展成为国内领先、国际先进的可扩展云化操作系统。目前 Cloudiip 已在 50 余个国家成功应用，接入 20 余类工业门类几十万台设备，汇聚全球数千余名活跃开发者，逐渐形成资源富集、开放共享、创新活跃、高效协同的平台生态，创造巨大的经济和社会效益。

基于 Cloudiip 工业互联网平台，东方国信为工业企业提供能源管理大数据系统，采用工业大数据、云计算、网络通讯、智能硬件等先进技术，实时采集工业现场的数字仪表、传感器数据，针对海量数据进行在线分析和挖掘，为企业能耗管理、设备效率提升、废弃物减少、生产计划安排等提供分析决策，助力企业进行能源管理对标和优化，可为企业带来 5%-15% 的节能效益。

1. 项目背景

联合利华是全球最大的日用消费品公司之一，公司成立于 19 世纪，在全球 100 个国家和地区拥有 163,000 名雇员，在六大洲拥有 264 个生产基地，产品涵盖食品、饮料、家庭及个人护理产品，旗下 14 个品类的 400 个品牌畅销全球 170 多个国家和地区。

联合利华不断追求在安全、高效、优质与环保等方面更优异的表现，其生产技术和设备非常先进，能源使用效率也位居世界前列。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成为一个庞大的系统，进一步的节能减排变得愈加困难，通过常规的技术改造难以挖掘节能潜力，利用大数据提升其能源管理水平成企业的重要课题。

2. 项目简介

东方国信基于工业互联网平台，为联合利华构建能源管理大数据系统，在联合利华全球 100 多个工厂进行实施，将各分厂的能源数据采集并传送到总部云平台，在总部云平台上对所有数据进行分析处理。

3. 项目目标

该项目旨在为联合利华建设能源管理大数据系统，具体目标有以下四点：

- 对企业能耗数据进行实时监测和可视化展示；
- 对能源生产、转换和消耗进行在线平衡；
- 以能源系统角度发现节能机会，挖掘节能潜力；
- 实现企业能源使用效率的提升，促进节能减排。

2、解决方案

1. 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 500 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，能源管理系统累计采集 50 多亿数据点，1500 多个活跃用户。

2. 技术方案

联合利华能源管理大数据系统基于东方国信工业互联网平台 Cloudiip，融合云计算、大数据、物联网等相关技术，包含数据采集、工业 IaaS 层、工业 PaaS 层以及工业 SaaS 层。



图1 东方国信能源管理大数据系统技术架构

● 数据采集及边缘计算

数据采集层通过不同协议及接口完成深层次数据采集并实现不同协议数据基层汇聚，作为工业互联网平台驱动源头。依托传感器、工业控制系统、物联网技术，面向设备、系统、产品、软件等要素数据开展实时采集。利用以智能网关为代表的新型边缘计算设备，实现智能传感器、设备数据和边缘分析结果云端汇集集成。

● 工业 IaaS 层

工业 IaaS 层为工业互联网平台提供所需的基础设施环境，包括存储、虚拟机、网络等，支撑资源的动态扩展和弹性分配服务，是工业互联网平台的重要保障及支撑。

● 工业 PaaS 层

工业 PaaS 层集成工业微服务、大数据服务、应用开发等功能，一方面将工

业领域积累的专业技术与知识经验进行模型化、组件化、微服务化，提供云化工业软件开发工具与应用环境，另一方面提供工业大数据系统环境，支撑海量大数据的接入、存储、分析和模型、共享，保障数据安全，形成完整的工业数据服务。

- 工业 SaaS 层

工业 SaaS 层面向工业能源管理场景，提供综合能耗分析与预测预警、机会识别量化与节能量化监测、故障预测和设备整体效率分析及能源专家系统等应用。

- 综合能耗分析与预测预警

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），建立合理的能效绩效目标来监控能源消费。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。

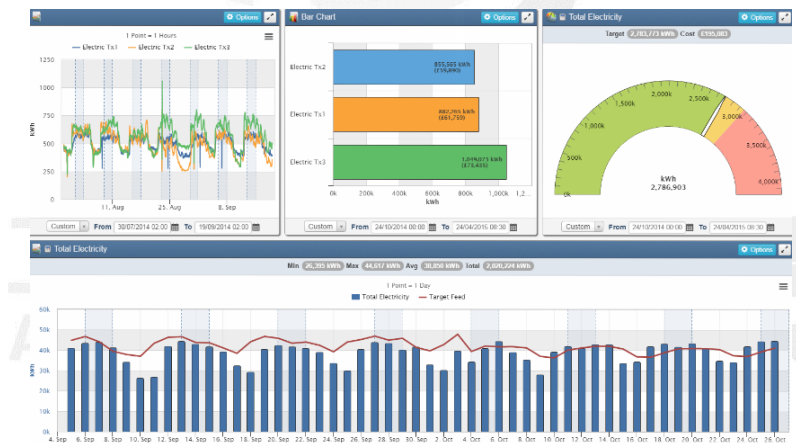


图2 能耗指标可视化监控

- 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。利用累计和图分析技术监测累计偏差，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。



图3 可视化分析能耗趋势

➤ 故障预测和设备整体效率分析

通过夹点、资产可用性及瀑布模型，开展故障预测和设备整体效率分析，使设备管理各环节得到系统性提升，为企业节约维修和停产费用。

➤ 能源专家系统

基于大数据分析，通过计算不同类别能源数据，建立涵盖两千多个方案的知识库，以此为基础形成锅炉专家系统、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。

- 1) 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。
- 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。
- 3) 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、凝结水回收率有效提升。
- 4) 制冷专家系统通过监控制冷环境并优化的冷却负荷，实现工厂可靠和高效方面的改进。将 PLC 系统与平台相连接，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。

3、实施效果与推广意义

1. 实施效果

截至目前，东方国信能源管理大数据系统已完成联合利华全球 100 多家工厂的能源数据的接入和分析，已接入的碳排放数据占全部排放的 50%，能耗数据占全部能耗的 54%，水耗数据占全部水耗的 64%。能源管理大数据系统的实施大大提高了联合利华的能源管理效率和水平，平均为每个工厂实现能源节约 5%-15%，节水 5%-30%，减少原材料 1%-3%，节约包装 5%，取得了巨大的经济效益和社会效益。

2. 推广意义

目前，我国工业企业的能源需求十分庞大，并存在能源利用效率低、结构不合理、供需不平衡、污染排放严重等问题，具有巨大的节能减排空间。能源管理大数据系统对推动“互联网+”智慧能源发展，推进能源监测、能量计量、调度运行和管理智能化体系建设，提高能源发展可持续自适应能力具有重要意义。

4、案例亮点

东方国信提供的能源管理大数据系统：

- 1) 采用了工业大数据、网络通讯、智能硬件、云存储等先进技术；
- 2) 实时采集工业现场的数字仪表、传感器数据；
- 3) 平台有更好的交互性和易操作性（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；
- 4) 数据采用云存储，磁盘容量无限制，可根据存储数量进行自动扩展；
- 5) 授权用户不受地域限制，也无需安装客户端软件，可以通过浏览器，输入网址，随时随地访问系统；

- 6) 平台可进行海量数据在线加工、统计、分析和挖掘，将能源与生产数据转化成易于管理和指导生产的简洁信息。

十、优也基础工业大数据平台在钢铁能效的应用

优也致力于运用工业互联网和工业大数据技术来提升基础工业的资源效率和运营绩效。优也根据具体的工业场景，以运营绩效的评估和诊断咨询为切入点，在能效管理，设备管理，排产优化和运营效率等方面，通过数据算法模型和工业互联网平台技术的深度耦合，开发整合成数字化智能化的专属解决方案，为基础工业的客户切实解决实际问题，提高资源效率，最终实现利润提升的目标。

1、案例背景与业务痛点

我国钢铁行业每年消耗 6 至 7 亿吨标准煤，约占全国煤炭消耗总量的两成，是我国节能降耗的重点行业。世界上主流钢铁企业多为钢铁联合企业，即囊括了从炼铁、炼钢到轧钢的多个工序。这样配置的一个重要原因就是有利于铁素（铁水、钢水、板坯）和煤气中能量的有效利用。其中，炼铁单元高炉、焦炉等设备产生的煤气用于轧钢等工序，可以让能源使用成本大幅降低、利用率显著上升。

由于钢铁联合企业的构成非常复杂，能效提升的推动过程会遇到各种困难。典型问题之一是煤气产生和消耗往往是不同步的：有时候产生多、消耗少；有时则产生得少、消耗得多，而每个工序都可能出现异常，带来整个管道压力的波动。如果不能保证有效的平衡，就可能造成无效放散或需要外购高价煤气以保证生产的正常运营。煤气波动过大时造成燃烧效率降低，即使已投用自动燃烧系统的炉窑也很难达到经济区间。煤气压力过低会造成之后工序因生产条件无法满足而临时停产。这样不仅会造成巨大的经济损失，也会造成环境污染。为了解决这个问题，需要有效地协调煤气跨工序的产生和使用。然而，要解决这个问题存在很多

困难：

1) 空间距离：涉及煤气生产和使用的包括高炉、焦炉、热风炉、转炉、连铸、加热炉、锅炉等非常多的设备。这些设备往往分布在钢厂的不同区域、距离常常相隔数公里，而且管路复杂，归属不同的生产管理部门。每一个工序主操都处于孤岛状态，只按本工序的需求来指导操作，导致总体协调控制困难。

2) 信息系统问题：数据来自于不同的设备和部门，形成信息孤岛，难以有效地集成和实时互连。煤气系统运行调配的决策依据与生产制造信息、工艺设备运行工况存在信息脱节，导致无法全面准确判断和给予精准决策。

3) 部门利益冲突：煤气相关的设备往往归属不同部门，各自的考核指标和利益诉求，难以从公司总体利益最大化角度优化煤气使用的配置。

4) 综合平衡问题：冶金行业的煤气系统种类繁多，都存在一个供需平衡，一个与产气装置、输配设施、使用设备紧密关联的动态的平衡系统。行业目前缺乏跨系统综合供需平衡及系统间平衡策略决策支持的软件应用。

5) 工况变化与异常的应对及时性：在这种大型生产系统中，不同工序都会遇到不可预料的异常状况，例如高炉修风，热风炉延迟换炉，锅炉负荷超载等等，而这种变化要求整个系统要对用气进行重新计算和分配，目前缺乏实时计算和调度功能。

6) 行业内已有的煤气信息化系统缺少人机交互的应用环境，系统使用者积累的宝贵的运行管理经验和调配策略知识无法有机植入系统，系统无法通过自我完善适应和满足不断变化的管理要求。

该案例针对东部地区某钢铁联合企业，该企业年产钢约 500 万吨，该企业的煤气系统在不同程度上存在上述的问题。各生产厂相距数公里，五台产气高炉数十个用气设备，各部门各设备之间的协同调度相当困难，生产过程中时有因煤气不足停产或煤气压力过高放散的问题，各部门的考核指标不尽一致甚至冲突，没有达成全网优化的统一策略。下面我们就以优也在该钢铁企业的高炉煤气智能平衡调度系统来介绍如何利用工业互联网和工业大数据技术解决钢铁行业的煤气实时寻优调配问题，感谢北京天泽智云在案例中提供的技术支持。

2、解决方案

1. 系统研发路线

对于高炉煤气系统平衡调配的智能化开发，需摸索高炉煤气自身系统的最佳运行优化方式、影响煤气使用制约因素等，更须考虑上述跨领域跨专业的结合和关联，从而实现高炉煤气系统内外兼顾的全方位智能联调运行。

通过对大数据的抽取归集与深度挖掘，建立一个基于经验关系、机理模型、数字模型等方法融合和关联的多维度信息库，以此形成高炉煤气系统信息价值链，将影响高炉煤气系统供需平衡的各类因素信息化、网络化、价值化，并贯穿平衡运行全过程。

2. 系统整体架构

如下图所示为高炉煤气智能平衡系统架构，系统分为 5 个架构层：设备层，资源层，数据与分析层，应用层和管理层。

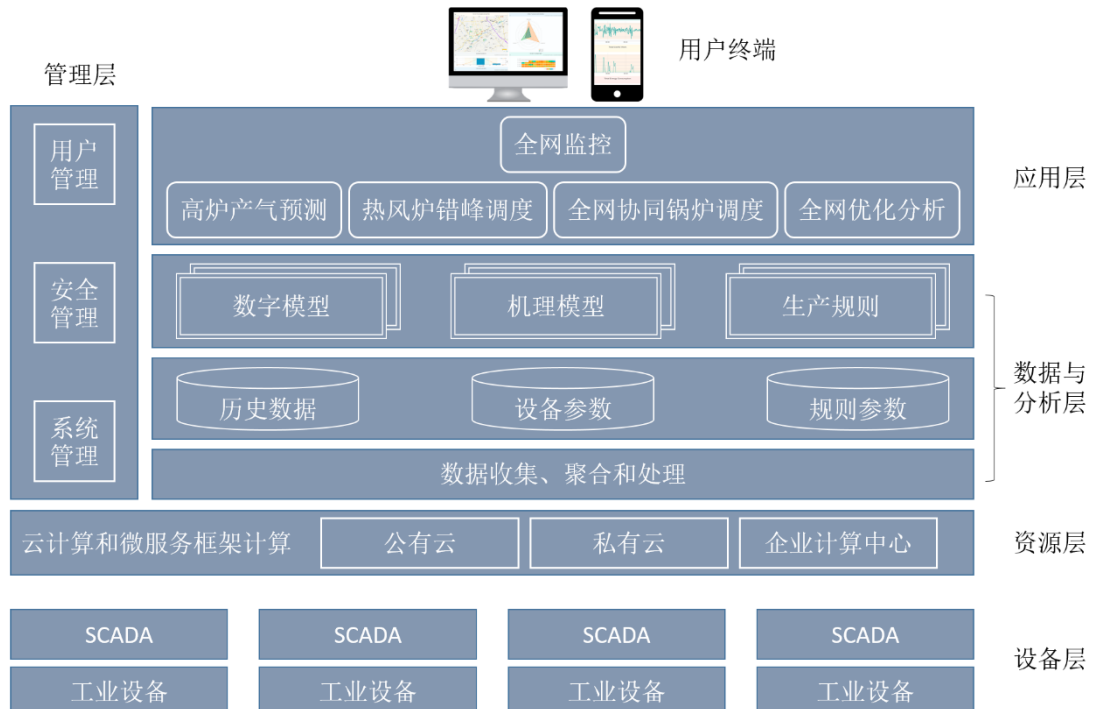


图1 高炉煤气智能平衡系统架构

1) **设备层**: 涵盖钢铁厂的设备，跨越多个工艺流程，一般由一个或多个 SCADA

系统连接和管理。

2) 资源层: 涵盖支撑本解决方案基于云计算技术和微服务框架的计算和网络资源。系统本身可以在共有云, 私有云以及企业内部的数据中心部署, 目前基于可靠性、时延以及安全性的考虑, 系统部署在企业内部的数据中心里。

3) 数据与分析层: 涵盖总体数据采集、处理与分析的功能, 以支撑上层的智能应用。

数据收集、聚合和处理的功能, 主要解决在生产过程中设备数据的多源异构、数据量大、并发性强、时延要求高以及一些数据质量参差不齐的问题。

数据储存的功能, 包括历史数据(经过过滤处理过的设备在生产过程中的状态数据, 用于建立和验证数字和机理模型)、设备参数(关于设备的规格、配置和设备之间的关系等数据)和生产过程的规则参数(用于配置生产过程规则的执行)。由于在生产过程中设备状态数据的采集量大, 而高质量数字模型的建立和验证也需要大量的历史数据, 历史数据的储存和分析需要采用大数据的技术框架。

支持数字模型、机理模型和生产规则模型的建立、验证和运行的功能。数字和机理模型主要用于对煤气的产用进行预测, 而机理和生产规则模型则用于根据预测的结果为各个工艺环节提出优化的煤气使用策略。模型建立和验证是以离线的方式进行。模型在经过严格的验证之后, 部署在系统上根据实时收集的数据提供在线的分析结果。

本解决方案从 SCADA 数据库里实时抽取与高炉煤气系统相关运行数据与产供用等各环节工况参数, 包括产气、用气、输配、贮存以及放散等关键数据。由于大部分冶金行业的 SCADA 数据相对完备, 所以在一般的情况下, 本方案的实施不需要另外加装传感器。但是原始数据也存在一些问题比如某些流量计不准, 无效数据, 数据延时等问题, 这些都需要在数据采集和处理过程中进行有效处理。

4) 应用层: 涵盖以数据分析的结果驱动的实现全网煤气调度优化决策的各个具体应用, 包括终端客户的使用界面, 提供针对具体工艺流程环节煤气使用的优化决策, 由生产人员操作执行, 系统对执行以及效果进行记录和反馈评估, 也提供整个系统在生产过程中的实时状态的可视化, 增加生产管理人员管理的透明度。

5) **管理层**：为整个系统提供系统管理、用户管理和安全管理的功能。未来保障系统的可扩展性、可靠性和可管理性，系统统一采用微服务框架实施内置系统服务和具体的客户端应用，并利用云计算技术对系统的计算资源包括数据、服务和应用进行统一的管理。

3. 面向用户的功能模块

高炉煤气智能平衡系统根据不同的用户日常操作和管理特点设置这 6 大功能，包括：

1) 对多个可能影响煤气平衡的因素实施信息跟踪和异常点分析，揭示即时的煤气产用的平衡动态。

2) 按“宏观-微观”、“浅显-深究”的方式，从几个维度“产供平衡”、“工序区域”、“单体机组”，运用统计分析工具，建立分析逻辑、经验计算等模型。

3) 通过对高炉未来产气量，预警波动并提前计算分配下游的用量并给出调节指导。

4) 从煤气保供、经济运行角度，梳理出相互关系，煤气与产线、其他能源介质系统的关联制定煤气调整的调度方式，覆盖煤气多种生成情况。

5) 整个管网策略里，将规则转化为专家系统，根据预先设定的安全限制与调度规则，自动指导缓冲调节。通过专家决策系统固化和执行运营调度规则。

6) 在不同模式下的经济核算和预估，以低成本运行和高利润为主旨，利用模型算法提出煤气调配策略下最佳成本路线，供用户交互式选择参数并测算该策略下的经济情况。

3、实施效果与推广意义

高炉煤气智能平衡调度系统在这家年产 500 万吨的钢厂上线以来，真正实现下面 5 个方面的显著效果：

1) 保生产: 系统上线前, 因为煤气压力过低造成轧钢的停线平均每月一次, 占总产量的 1%, 而系统上线后, 这种问题被彻底消除, 以吨钢净利润 500 元计算, 每年因提产可以带来 2500 万元的年化净利润收益。

2) 减放散: 煤气压力过高会造成放散带来能源损失, 系统上线后减少高炉煤气总量 3% 的放散, 以每立方米高炉煤气 0.1 元测算, 约合年化 1700 万的能源节降。

3) 降波动: 原来管网压力在选定区间的稳定率在 78% 的中位值左右, 系统上线后压力稳定率上升到了 95% 的中位值, 并且波动幅度极大缩小, 其结果是煤气使用效率的大幅提高和放散率的大幅降低。

4) 提效率: 系统通过模型化软件化将人的知识和经验传承下来, 指导一线生产人员进行标准化操作, 将操作从基于模糊经验化转变为科学化, 并为其提升提高了积累的基础, 提升决策敏捷性和精准性, 实现标准化操作的传承和提升。

5) 精管理: 系统自动预警产用气异常, 并对异常原因进行量化追踪, 辅助调度对煤气管网进行精细化管理, 实现了实时预警和异常追溯从无到有的突破。

该案例的原理有关如何运用工业互联网和工业大数据技术, 通过管理运营改善、运行信息关联、分析计算建模和工业互联网平台技术的深度耦合, 不仅适合钢铁行业, 还适合于有色、石油、化工等大型流程行业的燃气系统智能平衡与决策, 围绕燃气系统消除信息孤岛, 破除协同壁垒, 追踪运行动态, 识别异常事件, 辅助决策, 优化供需调配。

优也不仅用此原理和方法解决单一燃气系统优化决策问题, 也用于解决单一能源介质的多系统优化决策, 甚至是多能源介质的智能平衡优化决策。将此案例的原理和方法推广具有 3 方面的意义:

1) 直接价值: 给企业带来直接的效益回报, 无论是从降低波动, 提升产能这个角度, 还是从降低能耗这个角度都有直接可核算的经济效益。

2) 间接价值: 该系统降低了燃气压力波动, 为生产安全和环境保护提供了不可忽略的价值。

3) 社会价值: 燃气系统在基础工业里是非常普遍的一个生产系统, 试想如

果能在这些基础工业里都能够应用起来，对于国家的节能减排以及提升整体的工业效率将有深远的影响。

4、案例亮点

运用工业大数据技术最终目的是为了解决生产和经营中的实际问题，为企业切实地节省资源成本，带来利润地增长。本案例的核心亮点有下面 6 点：

1) 价值驱动与数据驱动：优也秉承顶级咨询公司的管理思路和方法论结合资深运营和产业专家的洞见，找到与企业的利润、成本和其它经济价值有关的关键创造点，并选取工序多，维度广，计算需求量大，很难通过人力能够做好的复杂场景来进行应用，解决关键问题来进行价值创造和管理提升

2) OT、IT 和 DT 深度耦合：同生产运营、调度策略等专家经验知识深度耦合，以生产规则和机理模型为根基，形成通过大数据建立的数字模型进行辅助决策，实时指导生产操作并形成反馈闭环，OT、IT 与 DT 相互融合与加强。

3) 基于工业互联网的实时优化决策：对于远距离，跨工序的全网协同才能实现时空范围优化、资源配置。

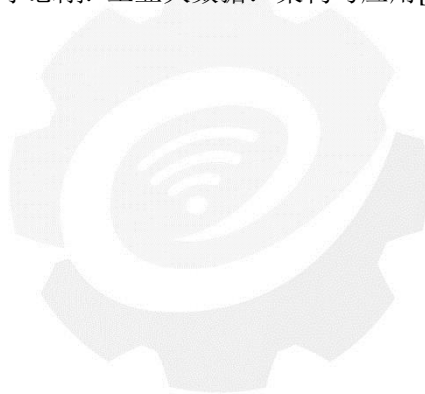
4) 工业大数据支撑：基础工业的数据条件相对完备，但是缺乏对数据的处理能力，运用工业大数据的采集，交换，处理，存储，建模，分析等技术能够实现不同设备，系统间的数据融合，最终可以基于数据提炼出来的信息，知识来进行优化决策。

5) 知识传承与智能化：传统的操作和调度知识经验存在于每个个体身上，无法进行共享和传承，通过对把行业中沉淀下来的知识和经验模型化算法化和软件化，将人的智能转移到系统的智能从而实现了知识的传承。

6) 管理水平提高：通过系统化重要指标分析和系统化优化策略入手形成一套自上而下的管理思路，不仅从管理思维上进行了梳理，更是通过智能化系统的固化将这种管理思维落实到日常的管理中，例如如何分析关键指标，如何追溯异常，如何快速响应和问题解决。

附录二：参考文献

- [1] 工业互联网产业联盟.中国工业大数据技术与应用白皮书[R].2017
- [2] 数据中心联盟.数据管理实践白皮书（1.0 版）[R].2017
- [3] 数据中心联盟.大数据平台选型与建设指南白皮书（1.0 版）[R].2017
- [4] 郑树泉, 宗宇伟, 董文生, 丁志刚. 工业大数据：架构与应用[M]. 上海:上海科学技术出版社, 2017



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

附录三： 缩略语

缩略词	英文全称	中文名称
AMQP	Advanced Message Queuing Protocol	高级消息队列协议
APS	Advanced Planning and Scheduling	高级计划与排程
CAD	Computer Aided Design	计算机辅助设计
CAE	Computer Aided Engineering	计算机辅助工程
CAM	computer Aided Manufacturing	计算机辅助制造
CAPP	Computer Aided Process Planning	计算机辅助工艺过程设计
CRM	Customer Relationship Management	客户关系管理
DDS	Direct Digital Synthesizer	直接数字式频率合成器
DCS	Distributed Control System	分布式控制系统
EMS	Environmental Management System	环境管理体系
ERP	Enterprise Resource Planning	企业资源计划
ETL	Extract-Transform-Load	抽取-转换-加载
IoT	Internet of things	物联网
JMS	Java Message Service	Java 消息服务
MES	Manufacturing Execution System	制造执行系统
MPP	Massively Parallel Processing	大规模并行处理

MQTT	Message Queuing Telemetry Transport	消息队列遥测传输
MSMQ	MicroSoft Message Queuing	微软消息队列
OA	Office Automation	办公自动化
PCS	Process Control System	过程控制系统
PDM	Product Data Management	产品数据管理
PLC	Programmable Logic Controller	可编程逻辑控制器
PLM	product lifecycle management	产品生命周期管理
RFID	Radio Frequency Identification	射频识别
SCADA	Supervisory Control And Data Acquisition	数据采集与监视控制系统
SCM	Supply Chain Management	供应链管理
SQL	Structured Query Language	结构化查询语言
WMS	Warehouse Management System	仓库管理系统
XMPP	Extensible Messaging and Presence Protocol	可扩展通讯和表示协议

法 律 声 明

本报告所载的材料和信息，包括但不限于文本、图片、数据、观点、建议，不构成法律建议，也不应替代律师意见。本报告所有材料或内容的知识产权归工业互联网产业联盟所有（注明是引自其他方的内容除外），并受法律保护。如需转载，需联系本联盟并获得授权许可。未经授权许可，任何人不得将报告的全部或部分内容以发布、转载、汇编、转让、出售等方式使用，不得将报告的全部或部分内容通过网络方式传播，不得在任何公开场合使用报告内相关描述及相关数据图表。违反上述声明者，本联盟将追究其相关法律责任。

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

工业互联网产业联盟 秘书处

联系电话：010-62305887

联系邮箱：aai@caict.ac.cn