

发电企业一体化 大数据平台白皮书

WHITE PAPER ON INTEGRATED BIG DATA PLATFORM
FOR POWER GENERATION ENTERPRISES

(2023)

联合发布



TRI 中安吉泰



中关村华电能源电力产业联盟
The China Chamber of Energy Industry

ABOUT US

关于我们

北京中安吉泰科技有限公司成立于 2015 年，作为能源行业智能运维领先者，公司以先进自研的高空作业爬壁机器人、数字孪生技术智能运维软件，以及新一代检测服务，专注为大型能源设施企业提供全栈式，集“监、诊、检、巡、安”于一体的全生命周期智能化运维服务。

公司产品业务覆盖火电、风电、核电、电网等领域，自主研发的 TRI 火电锅炉水冷壁爬壁检测机器人、TRI 火电锅炉智能运维四管健康分析系统等产品已广泛应用于国内各大能源企业。

公司目前拥有 49 项专利，40 项软件著作权，并主导了《电站锅炉水冷壁检修机器人通用技术标准》、《火力发电厂输煤廊道巡检机器人通用技术标准》2 项行业标准；参与制定了《工业机器人整机寿命加速试验性能退化评价办法》、《工业机器人动态稳定性评价指标及其试验方法》2 个机器人标准。

中安吉泰与武汉大学、华北电力大学、北京化工大学、北京邮电等高校长期保持战略合作伙伴关系。是国网、南网、中核、中广核、国电投、国家能源集团、华能、华电、大唐、长江三峡等集团的合格供应商。

中安吉泰在火电智能运维领域，已经在大唐王滩、鲁北，国能谏壁、泉州、华能莱芜、长兴，以及地方能源集团，如浙能凤台等几十家电厂部署了爬壁检测机器人，以及数字孪生智能运维系统；

在风电电站智能运维领域，也已在大唐赤峰、华能靖边、粤电阳江、石碑山等客户，部署了风电叶片、风电机舱、风电塔筒等检测清洗机器人及智能运维系统产品，在客户中赢得了极佳口碑。

两个第一

- 国内火电锅炉无人作业平台案例数量第一
- 国内首个针对风机实现全场景智能运维方案

标准制定

- 主导制定 2 个机器人标准：
《电站锅炉水冷壁检修机器人通用技术标准》
《火力发电厂输煤廊道巡检机器人通用技术标准》
- 参与制定 2 个机器人标准：
《工业机器人整机寿命加速试验性能退化评价办法》
《工业机器人动态稳定性评价指标及其试验方法》

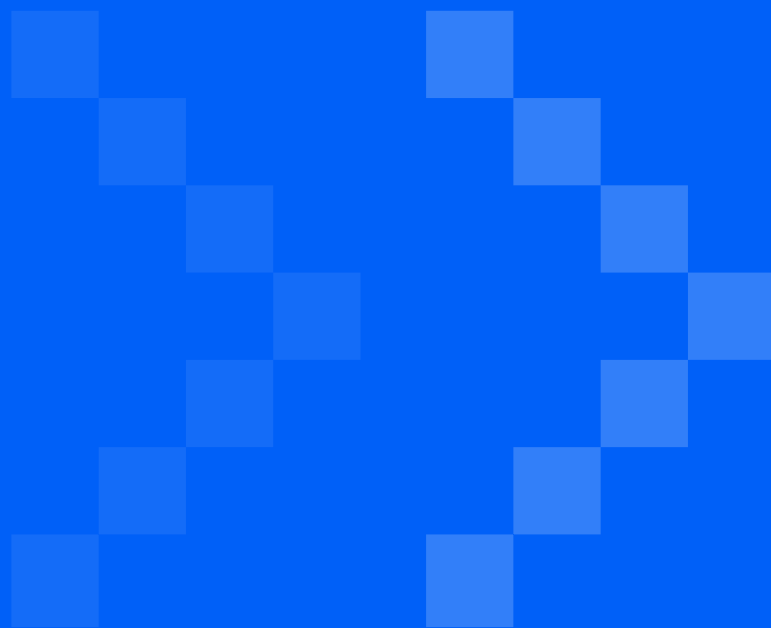
业界认可

- 2019 年“绿动未来”国家电投大数据及智能应用创新大赛三等奖
- 2020 年中国电机工程学会电力行业人工智能技术创新大赛二等奖
- 2021 年智能机器人系统通过中电联组织的“达到国际先进水平”的科技成果鉴定
- 2021 年能源发电智能机器人系统获得华为 ISV 认证
- 2021 年获得 CMMI3 软件能力国际认证
- 中核集团合格供应商认证证书
- 中广核 C2 合格供应商
- 挂轨机器人防爆认证证书



目录

CONTENTS



01

大数据平台 建设背景

BIG DATA PLATFORM
CONSTRUCTION BACKGROUND

.....06

02

一体化大数据 平台架构

INTEGRATED BIG DATA
PLATFORM ARCHITECTURE

.....06

03

一体化大数据 平台功能

INTEGRATED BIG DATA
PLATFORM FUNCTIONS

.....08

04

大数据平台 价值分析

BIG DATA PLATFORM
VALUE ANALYSIS

.....26

01 大数据平台 建设背景

BIG DATA PLATFORM
CONSTRUCTION BACKGROUND

智慧电厂建设是电力行业信息化发展的必然趋势，建立发电企业一体化大数据平台对电厂产生的各类数据进行统一管理是建设智慧电厂的关键环节，数据资产已经逐渐成为了发电企业的核心资产之一。随着电厂智能化手段的逐步提升，电厂各类信息化系统和数据量成倍增长，数据冗余、数据编码不统一、数据分散、数据闲置等问题成为发电企业普遍面临的问题，电厂庞大的数据未能得到有效利用。为了高效管理和使用这些数据，针对发电企业特点建设的一体化大数据平台应运而生，平台会将数据作为宝贵资产应用于设备监测预警、设备诊断、机组运行优化和企业经营决策等核心应用场景中，可以为发电企业的经营、生产、运行、安全和环保等领域的工作带来巨大的价值。

02 一体化大数据 平台架构

INTEGRATED BIG DATA
PLATFORM ARCHITECTURE

发电企业的一体化大数据平台是集电厂生产数据和管理数据的采集、治理、存储、计算和多维可视化展现等功能于一体的智慧基础设施中心。其技术架构图如下所示：





数据治理

数据治理是指将数据作为电厂核心资产而展开的一系列的具化工作，是对企业数据的全生命周期管理。数据治理是建立大数据平台的基础，其目标是实现数据在企业内部统一标准，提高数据的质量，保证数据的安全性、完整性及可用性，推进数据资源的整合、对接和共享。



数据采集

解决数据接入的问题，接收从软件系统或者硬件边缘端传输上来的数据，将数据采集到大数据平台中。为实现对多源异构数据的采集，需要支持多协议接入、数据预处理、实时规则、安全加密传输等能力。



数据储存

从电厂 DCS、SIS 和 MIS 等系统采集的数据一般分为三种形态，分别为传感时序数据（例如：外部环境信息，温湿度数据、压力流量、振动数据等），关系型数据（例如：设备 KKS 编码、设备检修台账等信息），文件数据（例如：存储视频、word 文档、图片等信息）。大数据存储平台能够实现对上述数据的集中存储，应具备分布式部署和横向扩容等特点。



数据分析层

数据分析层提供对数据的离线和实时计算能力，支持指标计算能力，人工智能分析计算能力和多维统计分析能力，可以为大数据的深入分析提供方法和工具。在数据分析层中通过实现对数据的深层分析挖掘，提供对经营分析、设备缺陷诊断、机组运行优化等方向的分析模型。



数据协作

数据协作提供了大数据服务接口、多维可视化展现服务等能力，支持将大数据平台中的原始数据以及分析计算后的数据以服务的形式对外部应用进行共享，同时实现对服务的集中管理和控制，减少重复开发工作，提高数据和模型复用率。



集中资源监控

通过集中资源监控平台对所有 IT 基础设施，包括硬资源（如主机、存储、路由器、交换机、防火墙和负载均衡等不同类型的网络设备）、基础软件平台（如操作系统、数据中台和中间件）、以及应用资源（如业务应用、网络应用和通信应用）进行全面统一监控和可视化管理，让用户可以实时了解、掌握网络硬件软件当前的健康状况，保障用户业务的可靠运行和满足各项考核指标，预测潜在的故障，进行提前预警。



统一认证管理

统一认证为企业应用提供集中的访问入口，实现高强度的多因素认证技术保障应用认证安全。建立集中、高强度的安全认证中心，以统一的安全认证策略和技术保障用户认证安全；通过定义数据中台各模块及业务应用接入规范和标准，实现当前主流的认证协议和认证技术，支持异构应用集成；通过上述技术，对人员、机构、权限等信息进行统一管理，解决新老系统的统一登录、认证问题及信息认证安全问题。

业务应用层

大数据平台可以有效支撑智慧电厂各领域应用的建设，包括但不限于：经营决策分析、生产运行管理、设备状态检修系统、机组运行优化系统、设备检修管理系统等内容。应用系统通过复用大数据平台中已经建设的相关能力，实现资源高效复用，减少系统建设时间和建设成本。

发电企业在建设一体化大数据平台时，建议将大数据平台的建设应与现有应用软件分开建设，确保其统一性、独立性、安全性和开放性。与应用系统之间即可独立部署又可有效复用，通过实现统一身份认证机制建立信任机制，通过服务接口或者功能模块链接的方式进行系统集成。

发电企业一体化大数据平台门户示意如下：



03 一体化大数据平台功能

INTEGRATED BIG DATA PLATFORM FUNCTIONS

- 1. 大数据治理平台
- 2. 大数据存储平台
- 3. 大数据分析平台

1.1 平台简介

数据治理平台涵盖了企业数据管理的全生命周期，包括：数据标准和编码管理、元数据管理、字典管理、数据建模、数据维护管理、数据集成管理、主数据管理、数据服务管理、数据质量管理、数据安全治理、数据生命周期管理等功能模块，打通数据管理各个环节，快速满足发电企业各类不同的数据管理要求。

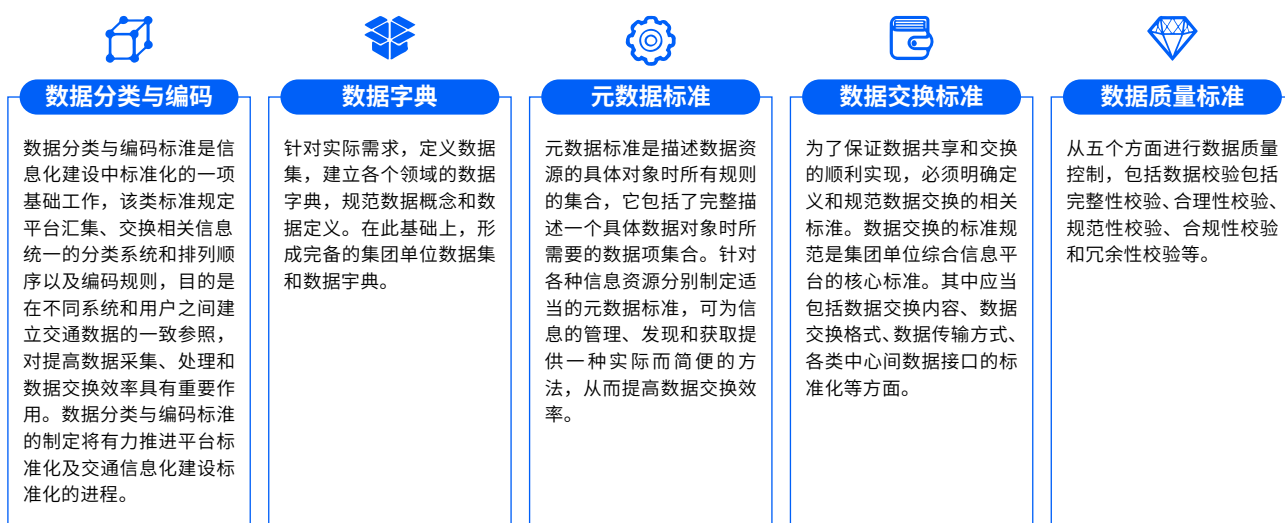
其功能架构图如下所示：



1.2 核心能力

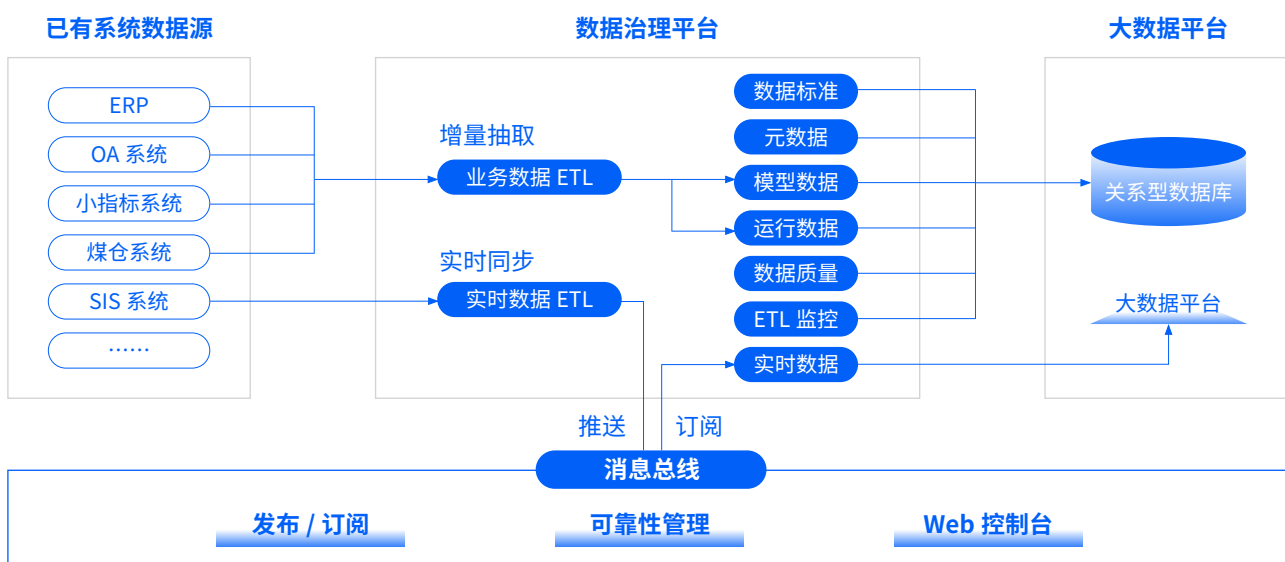
1.2.1 数据标准

数据治理对标准的需求可以划分为两类，即基础性标准和应用性标准。前者主要用于在不同系统间，形成信息的一致理解和统一的坐标参照系统，是信息汇集、交换以及应用的基础，包括数据分类与编码、数据字典、数字地图标准；后者是为平台功能发挥所涉及的各个环节，提供一定的标准规范，以保证信息的高效汇集和交换，包括元数据标准、数据交换技术规范、数据传输协议、数据质量标准等。



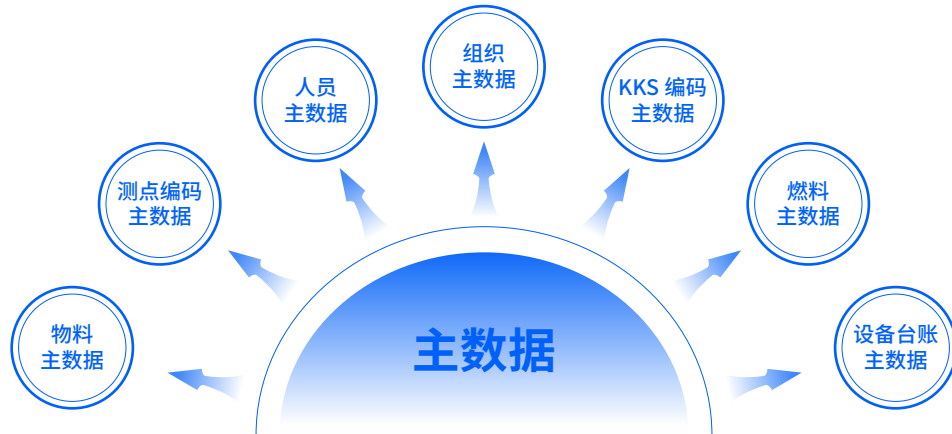
1.2.2 数据仓库 ETL

从现有业务系统数据源中抽取数据，经过 ETL 数据预处理后将数据存储到大数据平台中，大数据平台中的存储结构由元数据进行描述和定义，其中数据 ETL 过程包括：作业示例、作业模板、作业配置、调度时间、调度顺序、调度监控等如下功能。



1.2.3 企业级主数据管理

主数据管理是针对发电企业的特点定义核心业务模型，包括：设备、测点、机构、人员、物料等，然后从业务系统中将上述业务模型抽取出来形成主数据模型，为后继数据应用提供共享基础数据，减少各业务系统重复建设工作和因为主数据编码不统一导致的数据混乱。



1.2.4 提供数据服务

通过统一的数据服务平台来统一数据源，加快数据流转速度，提升数据服务的使用效率。将共享数据以服务的方式对外部系统进行共享，一般包括主动式调用（例如：RESTful 接口）和被动式接收（例如：Kafka）两种模式，提供服务前可以针对外部应用系统进行模型授权，指定哪些模型对那些应用进行数据共享。



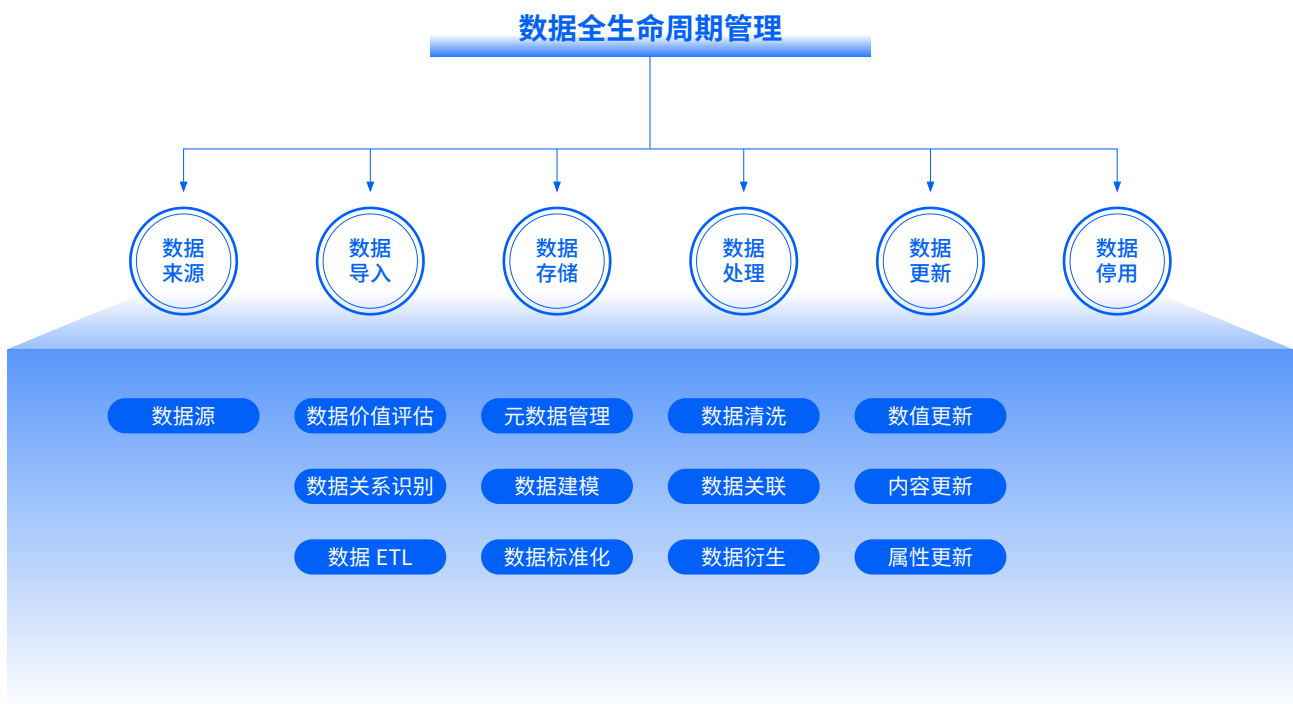
1.2.5 数据质量

数据质量模块支持数据校验包括完整性校验、合规性校验、规范性校验、冗余性校验和时效性五大模块，在设置校验规则时需选择校验类型、拥有者、错误等级、对象类型、校验日期等信息。

质量校验	基础数据	实时运行数据
完整性	分析数据库所有表属性的数据完整性，针对运行数据，要求数据是充分的，不存在缺失的记录和字段，属性值能够完整的描述所记录的业务	
合规性	保证数据唯一编码及拥有者符合结构化设计要求，基于结构化设计的规则库，针对数据库表属性取值、表内 / 表间属性比较、跨对象属性比较进行规则校验	
规范性	要求数据内容、格式符合通用校验表达式或业务规则表达式，以结构化设计作为校验依据	
冗余性	要求分析数据的记录是唯一的，不允许存在重复	
时效性	采集频率：分、时、天、周、月	评估数据的时间特性对业务应用的满足程度，采集频率一般是秒级

1.2.6 数据全生命周期

监控主数据和运行数据的全生命周期状态，包括数据的导入、存储、处理、数据更新和数据作废等状态，可以对数据变化情况进行全程监控。



03 一体化大数据平台功能

INTEGRATED BIG DATA
PLATFORM FUNCTIONS

1. 大数据治理平台
- 2. 大数据存储平台
3. 大数据分析平台



2.1 平台简介

发电企业的数据一般包括时间序列数据、关系型数据和非结构化文件数据等三种类型，存储方式分别为分布式文件存储数据库、关系性数据库和时间序列数据库。其中最复杂的是分布式时间序列数据库，因为时序数据具备传感器多、采集频率高、数据量大和计算模型复杂等特点，是发电企业中数据量最大的一种。

本文重点介绍时间序列数据的存储机制，这类数据在写入、存储和查询时的特点如下所示：

① 数据写入特点

写入平稳、持续、高并发和高吞吐，时序数据的写入是比较平稳的，这点与应用数据不同，应用数据通常与应用的访问量成正比，而应用的访问量通常存在波峰波谷。时序数据的产生通常是以一个固定的时间频率产生，不会受其他因素的制约，其数据生成的速度是相对比较平稳的。

写多读少，时序数据 80% 以上的操作都是写操作，是典型的写多读少的数据。这与其数据特性相关，例如监控数据，你的监控项可能很多，但是你真正去读的可能比较少，通常只会关心几个特定的关键指标或者在特定的场景下才会去读数据。

实时写入最近生成的数据，无更新、无删除，时序数据的写入是实时的，且每次写入都是最近生成的数据，这与其数据生成的特点相关，因为其数据生成是随着时间推进的，而新生成的数据会实时的进行写入。

无需传统数据库的事务处理。

② 数据存储特点

数据量大：以火电厂机组测点数据来举例，如果我们采集的测点数据的时间间隔是 1s，那一个监控项每天会产生 86,400 个数据点，若一台机组有 10000 个监控项，则一天就会产生 864,000,000 个数据点。

冷热分明：时序数据有非常典型的冷热特征，越是历史的数据，被查询和分析的概率越低。

具有时效性：时序数据具有时效性，数据通常会有一个保存周期，超过这个保存周期的数据可以认为是失效的，可以被回收。一方面是因为越是历史的数据，可利用的价值越低；另一方面是为了节省存储成本，低价值的可以被清理。

多精度数据存储：在查询的特点里提到时序数据出于存储成本和查询效率的考虑，会需要一个多精度的查询，同样也需要一个多精度数据的存储。

历史数据存储需要支持压缩机制，降低存储成本。

③ 数据查询和分析特征

- 按时间范围读取，显示一段时间内时序数据曲线。
- 最近的数据被读取的概率高，历史数据粗粒度查询的概率高。
- 单测点数据和多测点数据多种精度查询。
- 按照时间维度、区域维度和业务对象维度进行多维度查询分析。
- 有聚合计算需求。

时序数据需要用精心设计的时间序列数据库进行存储，区别于传统的关系型数据库和分布式文件数据库，时序数据库针对时间序列数据的存储、查询和展现都需要进行专门的优化，从而获得极高性能的数据写入能力、海量数据存储能力和极优的查询性能。



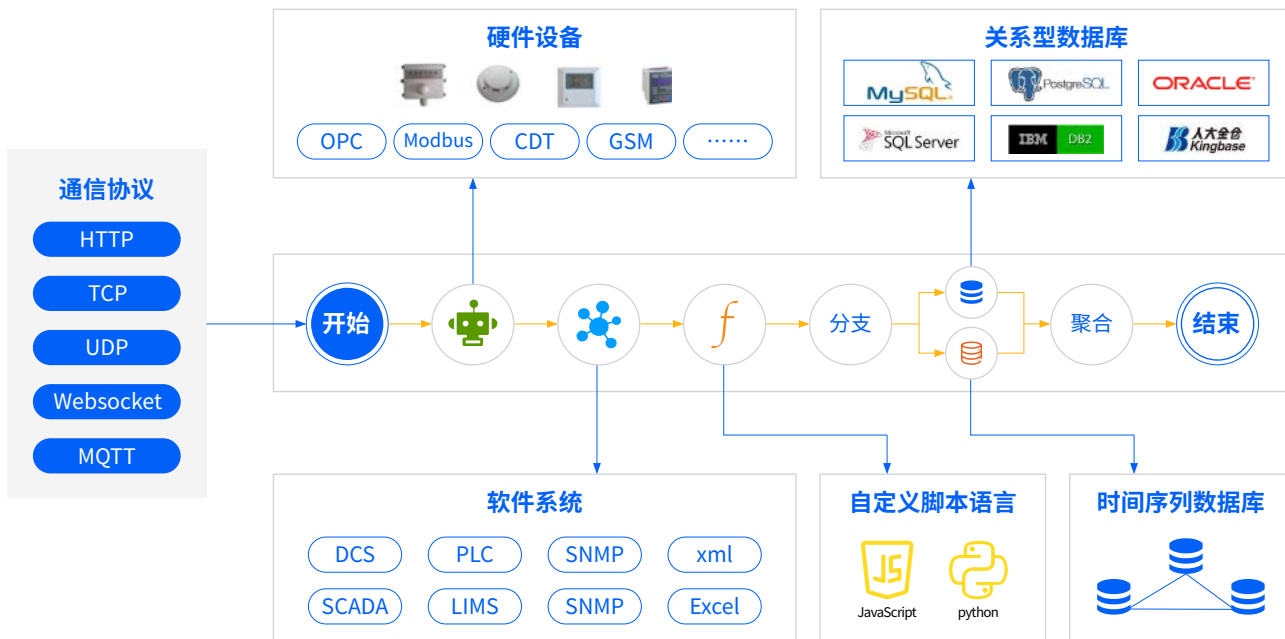
除了在数据写入、存储和查询等功能有其独特要求，时序数据的采集、分析挖掘和展现方式也与其他数据类型有明显不同，将时序数据连点成线，展示为时序图，可以用数据揭示已经发生的趋势性、规律性、异常性，借助机器学习和神经网络等技术可以实现对设备的故障缺陷预警和机组的运行优化分析。一般来说时序库要求具备如下功能：



2.2 核心能力

2.2.1 数据采集

数据采集模块采用的是可视化编程技术，通过拖动组件实现数据流处理，用于实时数据的采集，支持可视化配置式开发，支持拖动组件完成数据实时采集，函数组件支持常用业务开发基本的逻辑语法，满足网页编程，支持多种组件，能完成实时采集、实时计算和实时预警等功能。



2.2.2 时序数据建模

时序模型建模是根据业务需求定义模型结构，不同的数据格式，不同的时间戳格式，通过数据示例的结果转化，达到数据建模的目的。数据示例支持 Json 格式，特殊字符分隔符等。

	timestamp	unit_id	pi_code	original_name	quality	value
point	2019-01-20T19:25:21.046+08:00	1202268100000003	268130HSK31CF005	3_3HSK21CF003	0	31.09920883178711
	2019-01-20T19:25:21.046+08:00	1202268100000003	268130HSK31CF005	3_3HSK21CF003	0	31.09920883178711

timestamp

tag

field

point: 一个时序数据点，类似于关系型数据库中的 row;

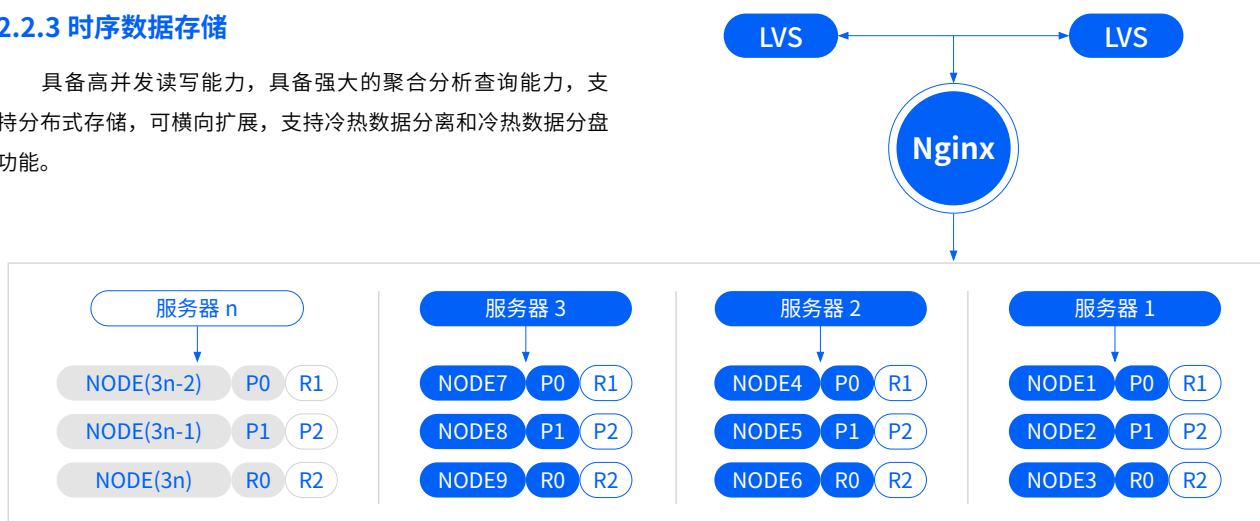
timestamp: 时间戳，表征时序数据产生的时间点；

tag: 维度列，用于描述设备 / 系统的属性，表明是哪个设备 / 模块产生的，一般不随着时间变化；

field: 指标列，用于描述设备 / 系统状态的变化，随时间平滑波动。

2.2.3 时序数据存储

具备高并发读写能力，具备强大的聚合分析查询能力，支持分布式存储，可横向扩展，支持冷热数据分离和冷热数据分盘功能。



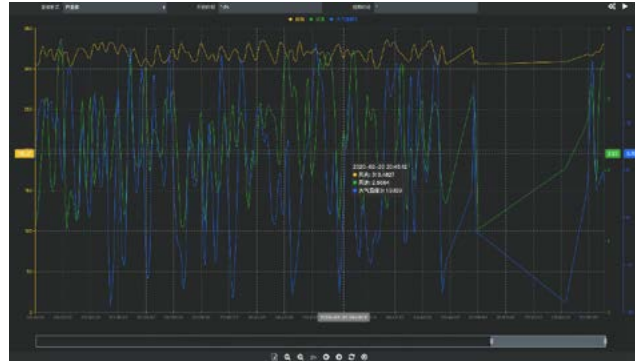
2.2.4 时序数据查询接口

为了方便业务应用访问时序数据库中的数据，时序数据的查询包括单测点数据查询、聚合计算查询、多测点查询等多种类型，下表提供了多种查询接口的定义格式，基本可以满足各类应用场景。

接口功能	接口描述	接口功能	接口描述
1 条件查询接口	核心基础接口，动态设定条件进行各类基础查询	8 批量获取某一时刻历史数据接口	查询历史某一时刻测点数据，当该时刻数据不存在时，会进行补点计算，返回计算后数据
2 聚合查询接口	核心基础接口，可以执行各类聚合查询	9 聚合计算接口	根据查询时间范围对测点数据进行聚合运算(包括: min、max、avg、sum)
3 单测点实时接口	根据测点名、系统编码和模型编码获取单一测点实时数据	10 历史实时接口	查询历史实时数据，根据查询时间返回距查询时间最近时间的数据
4 批量测点实时接口	根据测点名、系统编码和模型编码获取批量测点实时数据	11 加权平均计算接口	在查询时间范围内，按照数据所占时间比重计算该段时间数据的平均值
5 批量测点实时接口(s)	根据测点名、系统编码和模型编码获取批量测点实时数据	12 数据条数接口	根据条件获取一段时间内某测点的测点数据条数
6 批量获取历史数据接口 (小时段查询，建议不超过2小时)	根据查询时间范围，将时间进行分片，返回对应时间点数据，当数据不存在时，进行补点计算，并返回计算后数据	13 时序图查询接口	根据查询时间，等间隔返回测点(指标)结果
7 批量获取历史数据接口	根据查询时间范围，将时间进行分片，返回对应时间点数据，当数据不存在时，进行补点计算，并返回计算后数据	14 零点历史数据接口	批量获取零点的历史数据
		15 昨日历史数据接口	批量获取昨日的历史数据

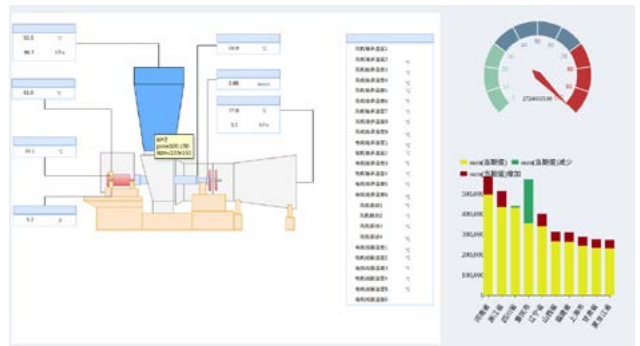
2.2.5 时序图

时序图分析 (Time-SeriesAnalysis) 是描述时序数据值在一段时间内变化波动的趋势和规律的统计工具，如整个数据值范围分布在什么范围内，具有波动较大的时期或时点等。通过对趋势的分析，也可以实现对未来数据变化的预测。



2.2.6 组态建模

组态建模是通过选择测点、指标等实时数据，实现组态画面实时数据的展示。并根据数据变换颜色、图形等形象的展示出设备或系统的运行情况，支持历史数据回放功能，支持动态添加右键菜单方式自定义显示页面，包括：数据看板、电子表格及查看时序图等。



2.2.7 数据告警

告警管理，系统内置六种告警规则，分别是：条件查询告警（简单）、条件查询告警（支持重复告警）、字段值统计告警、字段值连续统计告警、字段值分组统计告警、条件查询告警（数量倍数增长），支持六类场景条件告警。

配置告警规则后，当触发告警条件时进行提醒，系统内置邮箱提醒及微信，短信等提醒方式。



2.2.8 备份恢复

随着时间的变化，时序数据库占用的硬件资源越来越庞大，针对低频访问数据可以通过备份功能将数据迁移到低成本的硬件存储资源上，数据备份管理提供可视化备份管理界面，用户可通过选择系统，模型及时间来查询当前系统下的数据及进行数据备份操作。



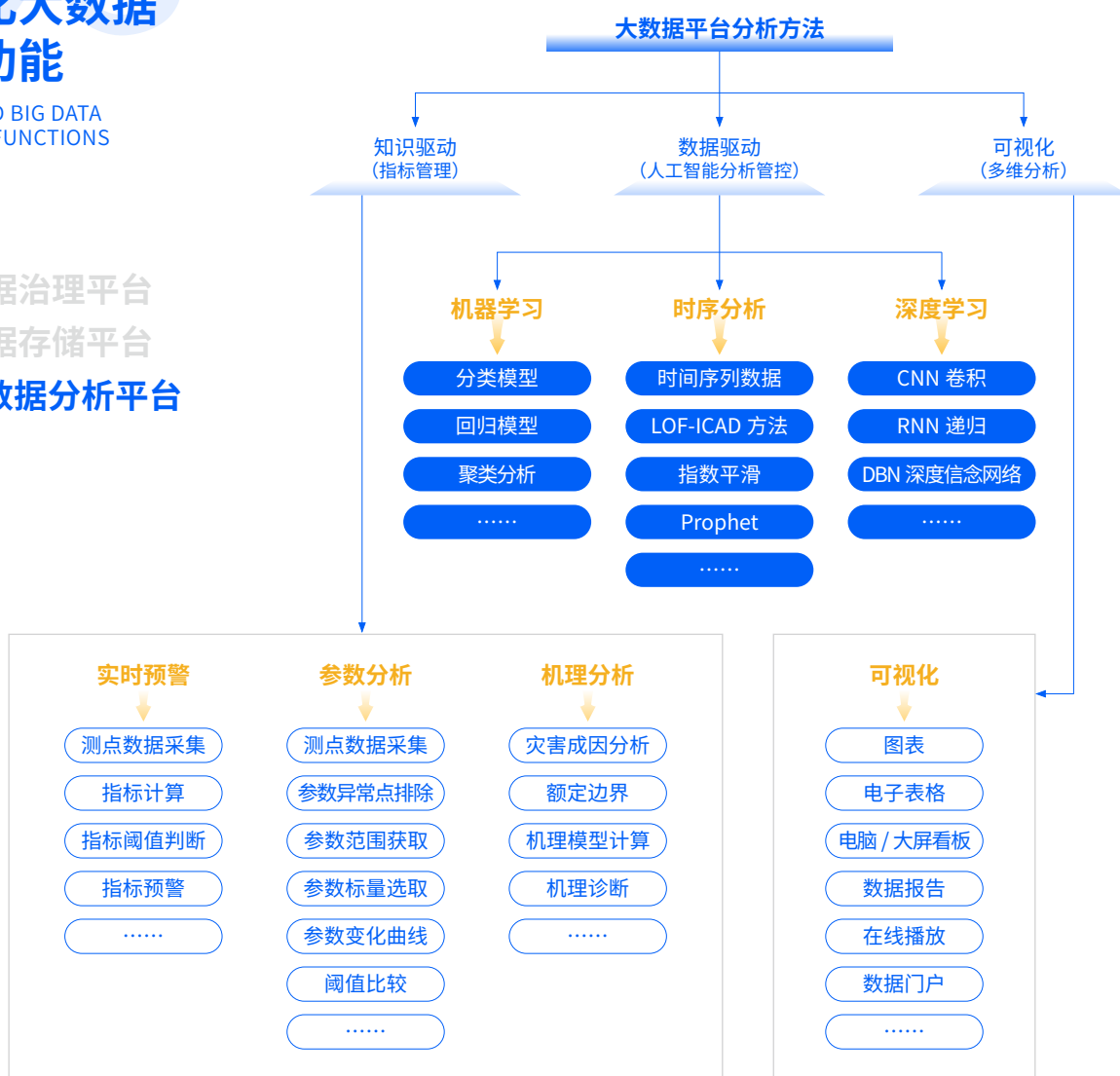
2.2.9 性能监控

用户查看时序库集群运行状态，可查看指标有 load（服务器负载）、process cpu（cpu 进程使用）、heap usage（集群节点内存使用）、disk usage（服务器磁盘占用）、uptime（解节点启动时长），通过上述功能可以及时了解时序库当前状态。

03 一体化大数据 平台功能

INTEGRATED BIG DATA
PLATFORM FUNCTIONS

1. 大数据治理平台
2. 大数据存储平台
- 3. 大数据分析平台



知识驱动的指标分析方法，是基于大量理论模型以及对现实工业系统的物理、化学、工艺以及行业经验等动态过程的知识，包括基于参数阈值、规则指标的方法和案例推理技术等，其中知识库是支撑这类方法的基础。

数据驱动的人工智能分析方法，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、聚类等方法。

基于统计学的数据可视化分析方法，主要是借助数据库的聚合计算能力，对数据进行分组求和、平均、最大值、最小值等计算，然后借助可视化工具对数据进行展示。

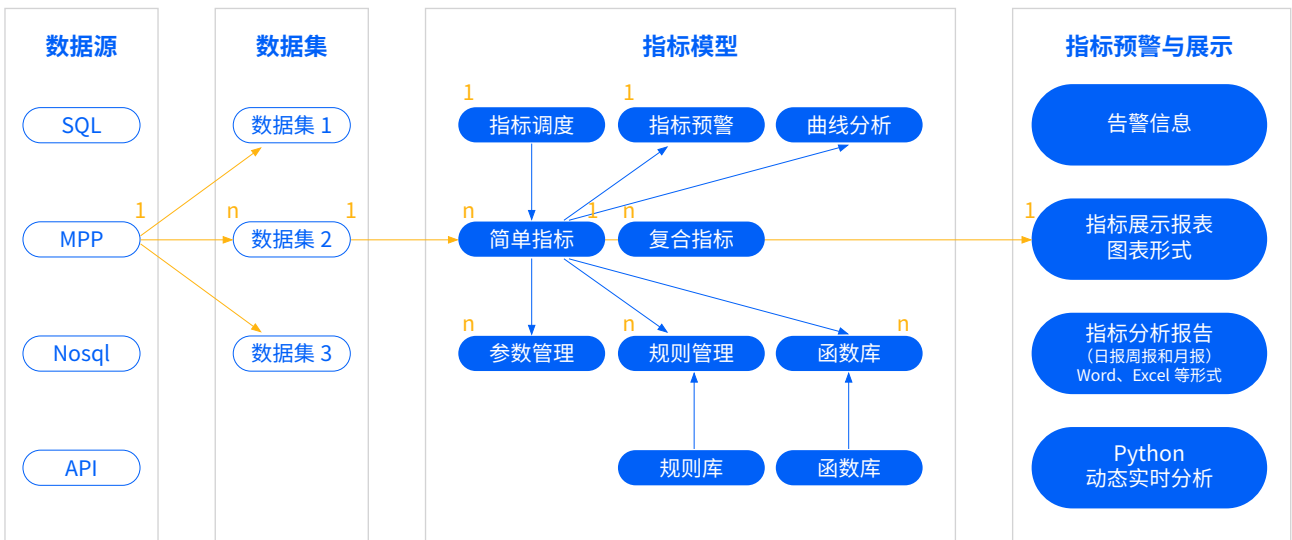
3.1 指标管理平台

3.1.1 平台简介

指标管理平台是一款基于大数据的指标在线计算平台。能同时面向业务人员和技术人员的对指标进行统一管理的工具。提供了数据建模、指标定义、指标数据落地和数据分析于一体的完整解决方案。



指标平台核心模型如下图所示：



参数：指标运算过程中使用的变量。

规则：指标的算法，通常一个规则包含多个参数。

调度：指标运算周期的控制。分为年、月、日、时、分、秒。

维度：指从不同的角度对指标数据的分类。分为时间维度、独立维度、多级维度、字典维度四种。

简单指标：简单指标指从数据集直接获取通过算法计算的指标，一般情况下，一个简单指标都会指定一个数据集和一个算法。

算法配置方式有四种，分为数据集 SQL、自定义 SQL、计算公式、自定义规则。

复合指标：复合指标是指在简单指标基础上，使用简单指标计算结果进行二次计算的指标。在高精度时间计算中（分、秒），为了满足计算时间的需求，会将复合指标还原至多个简单指标的计算规则的组合。

填报指标：填报指标是指由线下或无法通过数据集获取的数据，由人为手工填报的指标，可以用做后续的复合指标计算。

3.1.2 核心能力

指标管理平台应具备如下能力：

· **支持多种调度计算频率**，针对不同的业务场景合理使用不同的调度计算频率。如需要实时分析的场景数据，要使用分钟或秒级的调度频率来计算，系统将会根据算法将复杂指标还原至基本算法，进行统一计算，分发处理，顺利完成场景对计算时间的要求；如需要分析相对跨度大的场景数据，就要使用日、月、年的调度频率来计算，系统会根据复杂指标的层级，逐级执行，可以追溯各级指标对最终结果的影响。

指标计算中可以设置计算周期，计算周期分为：秒、分、时、日、周、月、季、年。



准实时计算特点：

1. 数据源查询参数数据块
2. 指标计算时效性高
3. 复合指标分解为原始计算公式
4. 在分布式环境下支持多节点分布式计算

准实时计算特点：

1. 逐级计算、指标计算结果复用
2. 手工填报
3. 指标计算结果手工修正
4. 指标重算

·支持多维度配置方式，针对相同算法指标使用维度配置方式，可以节省多余的指标配置数量，根据分析角度不同，通过配置不同的时间维度来进行精细化分析。



·支持算法函数，通过多个业务场景的实施，系统已将常用的算法封装成函数，以便于再次配置相同类型指标时直接使用。并且系统集成了 Groovy 脚本的配置方式，通过编写脚本的方式，实现任意复杂的算法，常用函数也可以在脚本中调用。



·支持便捷式定义和修改，系统直接通过页面配置指标，利用试算功能测试算法、结果精度等是否符合业务需求，在线调试和修改。

·支持对计算结果的预警、报警，通过设置指标结果的阈值来触发预警、报警功能，通过发送系统消息、邮件、短信等方式通知管理者及时做出相应的对策。

·支持重算、补录和修正。调度运行中如果出现数据源的数据传输错误或不可抗力的意外情况，恢复正常后可以通过重算功能来修正计算结果。如数据源端的数据并未恢复正常，也可以通过补录和修正的方式认为的修正结果。



3.1.3 适用场景

指标管理平台可用于如下业务场景：

① 经营分析统计

发电企业中对经营数据的分析是必不可少的，管理者需要通过对经营数据的分析做出阶段性的重要决策。指标管理平台通过配置指标的方式，可以针对管理者关注的经营指标进行定期计算，形成经营分析统计类指标，同时支持对分析统计指标进行拆解，通过追溯分析影响因素的构成，从而寻找出改善方向。

例如：在企业中，利润指标是所有管理者都非常关心的指标，利润指标一般是由收入指标减去成本指标，收入指标又由主营业务收入和其他业务收入构成，依次追溯下去，我们会从企业最基本的财务数据中获取数据，然后计算出利润指标，从这个指标计算过程，我们能够帮助管理者找到问题点，进而提升企业经营能力。

2 指标监控告警

在能源行业，无论是电厂的机组还是电网中的变电站都有大量的实时指标数据在跟踪当下运行情况，一旦出现问题，就会影响到电厂正常发电和变电站正常运行，造成重大经济损失。指标管理平台通过对指标的实时计算监控，针对相同算法的指标，通过维度的配置方式，可以节省指标数量的配置，并设置阈值预警和报警，提高安全系数，提早做出预防工作，避免损失。

3 设备故障诊断指标场景

发电企业设备一般大多具备价值高、运行复杂等特点，要想准确的诊断甚至预测设备的故障，难度很高，因为设备的核心参数之间不是彼此孤立的，会相互影响，因此基于机器运行机理建立多参数之间的关联分析指标就非常有必要，我们会将核心关键参数根据设备机理形成指标计算公式和特征曲线，借此方法分析判断设备异常情况。

同时，当指标模型建立后，可以对历史指标结果数据的故障因素进行分析，形成指标故障模型，待某些指标值达到某个特定条件时，系统会自动分辨预测故障状态，此时可以根据报警，进行故障方向的排查，减少故障的发生。

4 工况寻优类指标计算

在发电企业中，影响设备在最优的工况下运行的因素有很多，总来的来说分为两大类，一类是可控因素一类是不可控因素，当管理者想要寻找最优的方案时，不知从何下手。指标管理平台通过对同一参数不同数值的监控计算，寻找到最优的参数，提示设备操作者将可控因素进行调整。同理，还可以通过历史指标数据的分析，寻找历史最优值。

例如：工况优化中发电厂中对安全和效率同时关注，而温度、压力等是可调整的因素，而效率又是不可控的，可以通过同指标不同算法的配置，寻找出相对安全，效率又最高的临界点。

3.2 人工智能分析管控平台

3.2.1 平台简介

随着计算能力越来越强，云计算、大数据、虚拟化等技术的出现，让人工智能有了可依赖的现实技术基础。随着大数据的积聚、理论算法的革新、网络设施的演进，机器学习进入全新的发展阶段。人工智能的算法需要依赖海量的数据，利用海量的样本进行机器学习，利用大数据技术分析，帮助企业深度数据挖掘、分析数据、使用数据。

机器学习算法平台基于微服务结构，依托机器学习和深度学习引擎，将数据输入输出、数据预处理、挖掘建模、模型评估等环节通过在线流程化的方式进行连接，达到数据分析和数据挖掘的目的。



3.2.2 核心能力

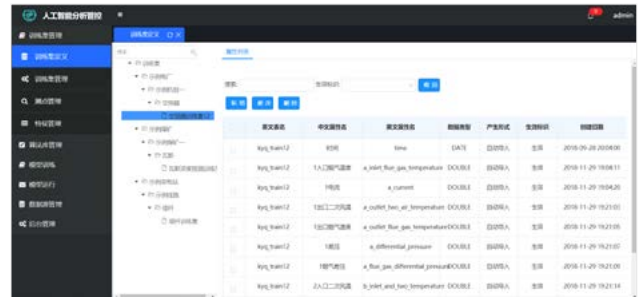
人工智能分析管控平台应具备如下能力：

① 训练集模型元数据管理

支持动态添加训练集模型，定义训练集模型表的字段名称、字段类型、长度等信息。在模型训练过程中需通过关联分析方法不断完善，因此需要此功能在浏览器上进行操作，可动态扩展。

② 训练集管理

训练集模型数据增、删、查改信息管理，训练集数据的选取极为关键，需要借助专业知识从现有历史数据中进行挖掘和发现，有效训练集数据越多、正负样本数量越完整，模型会越精准。



模型名称	模型描述	模型类型	训练集	测试集	创建时间	更新时间
Model1	1号炉气温度	a_hot_hot_gas_temperature	OC30.0	OC30.0	2018-09-28 09:04:00	
Model2	炉温	a_hot_hot_gas_temperature	OC30.0	OC30.0	2018-11-29 18:08:11	
Model3	炉温	a_hot_hot_gas_temperature	OC30.0	OC30.0	2018-11-29 18:08:26	
Model4	炉温	a_hot_hot_gas_temperature	OC30.0	OC30.0	2018-11-29 18:21:01	
Model5	炉温	a_hot_hot_gas_temperature	OC30.0	OC30.0	2018-11-29 18:21:05	
Model6	炉温	a_hot_hot_gas_temperature	OC30.0	OC30.0	2018-11-29 18:21:07	
Model7	炉温	a_hot_hot_gas_temperature	OC30.0	OC30.0	2018-11-29 18:21:09	
Model8	炉温	a_hot_hot_gas_temperature	OC30.0	OC30.0	2018-11-29 18:21:14	

③ 算法库管理

该模块主要功能是建立一系列的算法组件，支持算法的添加、修改和删除功能，在算法库中封装发电企业常用的算法模型，包括：数据预处理算法、分类算法、回归类算法、图像识别类算法、时序分析算法等。

④ 模型训练管理

基于任务流方式配置人工智能模型训练过程主要包括 6 个主要步骤：

① **确定数据源**，指定上述训练集对象作为机器学习模型的数据源。

② **数据预处理**，数据集中可能包含各种数据异常情况，例如：数据值缺失、数据值错误、数据类型异常等等，这些情况都会导致损害模型的准确性，通过对数据进行预处理，来消除数据中的噪声、纠正数据的不一致、识别和删除离群数据、对数据进行规约管理，来提高模型鲁棒性，防止模型过拟合。

③ **特征工程**，用数学转换的方法将原始输入数据转换为机器学习模型的新特征，或者借助于领域专业知识应用于机器学习模型，挖掘和发现新的特征，这些使得特征能在机器学习算法上发挥更好的作用的过程，包括：特征初步筛选、特征变换、特征组合、特征选择、数据的归一化处理等等。

④ **算法选择**，根据业务场景选择算法，包括分类、回归、聚类、时序分析算法等等。

⑤ **模型运行**，设定训练集与测试集比例，进行模型评估，计算出模型评测结果。

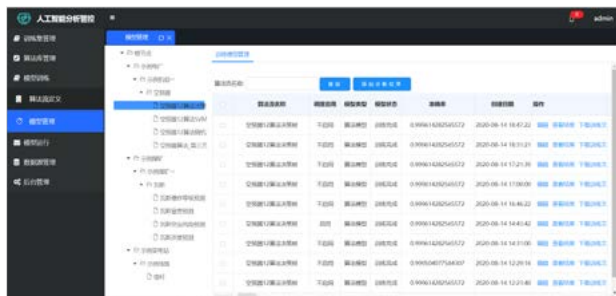
⑥ **模型存储**，当模型预测准确性达到预定标准后，将模型进行持久化存储。

模型训练流程图中定义模型训练的全过程，包括特征工程，数据预处理，数据分类，核心算法，准确率，分类报告等信息。具体如下图所示：



5 模型管理

对训练后的模型进行统一管理，包括：模型添加、版本管理、模型下载、模型作废等等。



6 模型运行配置

设置模型调度执行周期、模型运行实时数据来源等信息。

7 模型诊断结果确认

针对人工智能模型判别结果进行人工确认，如果确认判断正确，人为对此结果进行确认，同时将判别对应的数据集动态添加到训练集中，为下一版本的模型训练增加训练集数据，对模型进行优化。

3.2.3 适用场景

人工智能分析管控平台可以用于以下业务场景分析：

1 设备故障诊断

在电厂中，存在大量转动设备，传统的诊断方法是基于设备参数超限预警和设备机理的分析，在智慧电厂中，积累了TB级的设备运行数据，可以依托大数据平台，使用基于机器学习和深度学习技术，通过对接设备故障库、算法、构建模型、迭代优化等步骤逐步完善设备故障智能诊断预警系统。

2 锅炉四管健康分析

通过人工智能的深度学习模型对“四管”历史环境测点数据和减薄测点进行筛选，处理和模型优化，然后利用线上实时的测点数据可以对一定周期内的“四管”减薄进行预测，在“四管”经过一段时间减薄之后可以对“四管”进行防爆预警。要求对历史的减薄数据进行对比测试，此外能够提供“四管”防爆预警模型训练平台以及线上部署服务的平台。

3 机组运行优化场景

在机组运行优化系统中，采用人工智能算法与模型将当前工况与历史工况进行聚类分析，计算得到在此工况下历史上的最优运行参数值，从而指导当前工况下机组运行参数的调整。

4 输煤廊道巡检场景

在输煤廊道巡检场景中包括：皮带机上异物闯入监测及告警模型，皮带机侧方撒煤监测及方位告警模型，人员跌倒、未佩戴安全帽等安全分析模型。

3.3 多维可视化分析平台

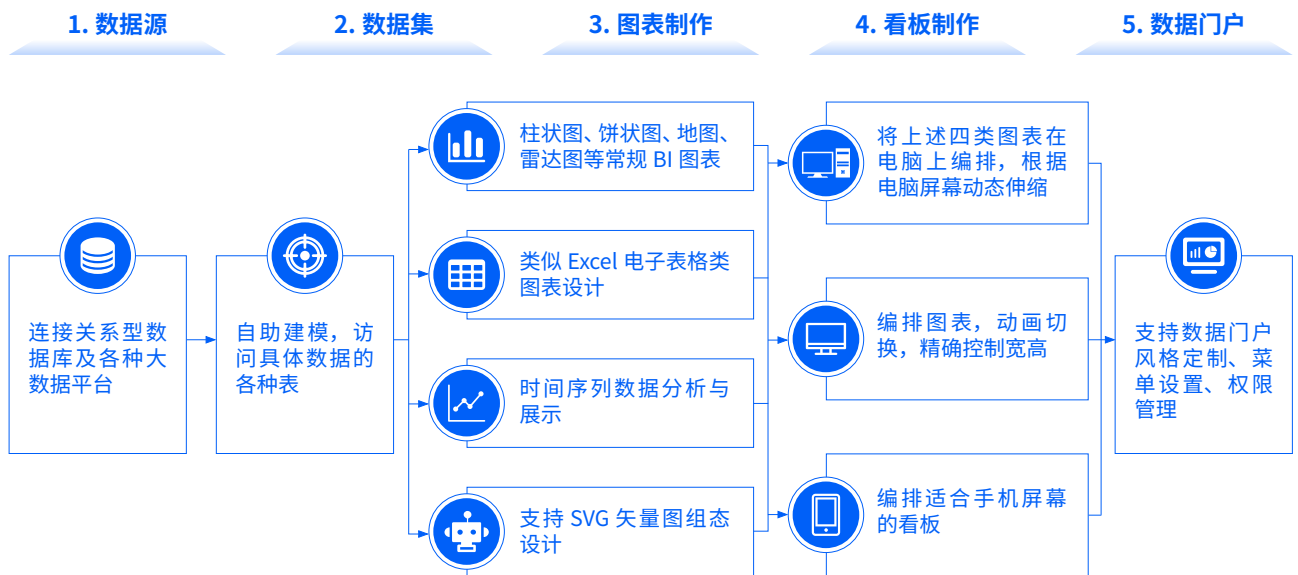
3.3.1 平台简介

多维可视化分析平台是一款定位于工业互联网业务场景下的数据分析和可视化展现的产品。通过数据源的连接和数据集的创建，对数据进行分析 and 查询；通过丰富的图表组件和看板组件以及电子表格组件，以拖拽的方式进行数据可视化的展现。

多维可视化分析平台能够实现可视化报表业务灵活方便的扩展，不仅能对具体报表业务进行调整，还能灵活方便增加新的报表业务，实现企业统一报表口径，提高数据质量，提升业务人员和管理者的数据获取效率。



多维数据分析平台开发过程如下：



3.3.2 核心能力

多维分析可视化平台应包括如下能力：

① 数据源管理

数据源管理用于定义数据来源，支持目前主流的关系型数据库，如 Oracle、SQLServer、DB2、MYSQL、PostgreSQL 等，以及常用的国产数据库，如达梦、人大金仓、GBase 等；大数据架构下的数据库支持 Hive、ElasticSearch、Kylin 等；另外系统还实现对文件导入数据、RESTAPI 读取到的数据进行分析、展示。同时可通过扩展的方式添加新的数据源到系统中。

② 图表组件

支持丰富的图表组件，支持过滤类组件、2D 图表组件和 3D 图表组件等。

· 过滤组件

日期过滤组件、列表过滤组件、下拉框过滤组件、树过滤组件、LookUp 查询过滤组件等。

· 2D 图表组件

支持按钮组件、饼图组件、区域图组件、柱状图组件、曲线组件、水球图、漏斗图、桑基图、仪表盘组件、柱状图组件、散点图组件、环形图组件、雷达图、时间迁徙图、GIS 组件、度量组件、交叉表格组件、时序图组件、直方图组件等。

· 3D 组件

支持三维饼状图组件、三维区域组件、三柱状组件、三维曲线组件、3D 环形图组件、堆积柱状图组件、三维散点图组件等。

③ 电子表格

电子报表设计器是用于解决复杂类似 Excel 表格类业务场景，通过迭代单元格可以实现任意复杂的中国式报表，设计器可以在 Chrome、Firefox、Edge 等各种主流 Html5 浏览器运行。使用电子报表设计器，打开浏览器即可完成各种复杂报表的设计制作。而且报表支持在线打印。支持类 Excel 设计模式：兼容 Excel 公式，支持多 sheet，支持单元格无限扩展，支持复杂报表的设计。

安全运行3541天															
日期:	2018年7月31日														
指标	发电量	弃电量	上网电量	厂用电率	上网电价	上网电度	上网电量	上网电量	上网电量	上网电量	上网电量	上网电量	上网电量	上网电量	上网电量
单位	万千瓦时	万千瓦时	万千瓦时	%	元/千瓦时	万千瓦时	万千瓦时	万千瓦时	万千瓦时	万千瓦时	万千瓦时	万千瓦时	万千瓦时	万千瓦时	万千瓦时
计划值	10000	8000	8000		0.100000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000
实际值	10276.88	8282.02	8282.02	8.08	0.100000	10276.88	10276.88	10276.88	10276.88	10276.88	10276.88	10276.88	10276.88	10276.88	10276.88
完成率	102.77%	103.40%	103.40%	8.08%	0.100000	102.77%	102.77%	102.77%	102.77%	102.77%	102.77%	102.77%	102.77%	102.77%	102.77%
完成比例	102.77%	103.40%	103.40%	8.08%	0.100000	102.77%	102.77%	102.77%	102.77%	102.77%	102.77%	102.77%	102.77%	102.77%	102.77%

④ 看板设计

看板设计可根据业务需求将业务人员关注的报表灵活组合在一个主题页面上，支持在大屏端、电脑端和手机端进行展示。在主题看板上，用户可以随时通过筛选器对数据进行过滤，可以通过图间交互联动实现多维度数据关联分析，可以通过层层下钻进行数据透视分析从宏观到微观追溯原因，可以通过跳转功能将多个仪表组合成一个更大的分析主题，可以通过维度的旋转对同一个指标进行多角度的查看与分析，便于用户全面掌握、多方位深层次分析、从宏观到明细层层追溯和多业务指标关联分析。



⑤ 数据报告

数据报告生成功能主要用于定制化图文并茂的报告自动生成。通过定制化设计好各种报表的模板，并且对接数据源或数据仓库，只需一次设计后续图文并茂的报告将自动生成，报告中的文字、数字、图表等根据当前最新数据动态生成最新报告。



3.3.3 适用场景

多维可视化分析平台可用于如下业务场景：

① 多维数据业务分析场景

发电企业业务人员和技术人员有对业务数据进行多维度、可视化分析的需求，例如：在企业经营决策系统中，需要开发大量图表和看板，此时可以基于多维数据分析平台进行快速配置开发，搭建多维分析仪表盘，进行各类业务指标和数据分析展示。

② 电子表格分析场景

发电企业中，有大量固定格式化的基于 Excel 设计的报表，例如：生产日报、节能月报、企业年度财报等等，此时可以借助电子表格设计模块进行开发，整个开发过程与 Excel 操作相似，支持现有 Excel 模板格式导入，无需专门培训，可快速上手，快速实现与企业现有报表样式完全相同的在线报表开发。

③ 大屏展示场景

在发电企业中大屏展示场景越来越多，通过大屏集中监控，包括：指挥调度、业绩展示、运维监控等等，大屏展示要求与电脑展示存在一定的差异，包括能够根据大屏屏幕尺寸精确控制图表显示位置、能够动态效果轮播，动态效果展示等等。

④ 移动展示场景

定制移动端数据看板，便于发电企业管理者随时随地查看最新的企业经营、运行数据，为决策提供数据支撑。



建立发电企业统一的数据标准和数据编码体系，发电企业机组、设备、传感器、物资、备品备件、设施种类繁多，分布范围广，设备管理过程中涉及大量的实时数据、巡检数据、检修数据、技术资料、设备图纸和档案等数据类型，大数据平台可以助力电厂建立一套能够贯穿电厂全生命周期整个过程的唯一标识，将上述设备与数据进行关联，实现统一管理，便于不同时间阶段、不同专业、不同厂家的各类专业人士准确理解和交流。

建立发电企业统一的数据治理标准、统一的数据存储机制和统一的数据模型，并针对不同的数据模型制定统一的数据采集策略和数据质量校验规则，保证数据采集的实时性、准确性和唯一性，确保多源异构数据的统一高效安全存储，解决目前发电企业普遍存在的数据孤岛问题。

建立统一的数据服务共享能力，为智慧电厂中的各业务应用模块提供数据复用的基础。减少重复数据采集成本，减少跨专业反复沟通协调成本，提高数据使用效率。

建立统一的数据计算分析挖掘模型，针对发电企业的特点建立业务指标模型中心和人工智能模型中心和报表展示模型中心，当电厂专工对上述分析模型结果校对确认后，即可发布到三大模型中心中，在多业务应用系统中复用上述模型，达到一次建立模型重复使用模型的效果。

TRI 中安吉泰



📞 17801031890

☎ 010-84463178

🌐 www.3h-technology.com

📍 北京市海淀区创业路 8 号 3 号楼东侧 3 层 3-4 室 301

