目录

[1 项目介绍 4](#_Toc501540811)

[2 项目总体设计 7](#_Toc501540812)

[2.1 设计概述 7](#_Toc501540813)

[2.2 项目目标 9](#_Toc501540814)

[2.2.1 成本精益控制 9](#_Toc501540815)

[2.2.2 质量精益管控 10](#_Toc501540816)

[2.2.3 客户精益服务 10](#_Toc501540817)

[2.2.4 设备精益监控 10](#_Toc501540818)

[2.2.5 安全精益监督 11](#_Toc501540819)

[3 系统框架 12](#_Toc501540820)

[3.1 系统框架总体结构 12](#_Toc501540821)

[3.1.1 总体逻辑框架 12](#_Toc501540822)

[3.1.2 总体技术框架 13](#_Toc501540823)

[3.2 数据获取 15](#_Toc501540824)

[3.2.1 实时镜像备份 15](#_Toc501540825)

[3.2.2 网络爬虫scrapy 15](#_Toc501540826)

[3.3 数据仓库 17](#_Toc501540827)

[3.3.1 数据仓库设计 17](#_Toc501540828)

[3.3.2 数据仓库管理 17](#_Toc501540829)

[3.3.3 数据接入 18](#_Toc501540830)

[3.3.4 数据清洗 20](#_Toc501540831)

[3.4 知识挖掘和模型构建 22](#_Toc501540832)

[3.5 应用服务与智能平台 24](#_Toc501540833)

[3.5.1 应用服务后台 24](#_Toc501540834)

[3.5.2 智能可视化平台 25](#_Toc501540835)

[3.6 数据服务平台 26](#_Toc501540836)

[3.7 系统信息安全 28](#_Toc501540837)

[3.7.1 基于CPK数字签名技术的身份认证方案 29](#_Toc501540838)

[3.7.2 基于CPK数字签名技术的登陆记录不可抵赖方案 30](#_Toc501540839)

[3.7.3 基于CPK的文件加密解密方案 30](#_Toc501540840)

[3.7.4 基于CPK的安全数据库方案 30](#_Toc501540841)

[3.8 消息推送服务平台 31](#_Toc501540842)

[4 大数据与系统集成 33](#_Toc501540843)

[4.1 工业大数据平台 33](#_Toc501540844)

[4.2 Spark分布式计算框架 34](#_Toc501540845)

[4.2.1 Spark概述 34](#_Toc501540846)

[4.2.2 MLlib数据挖掘算法库 35](#_Toc501540847)

[4.3 Hadoop生态系统 35](#_Toc501540848)

[4.3.1 HDFS(Hadoop分布式文件系统) 35](#_Toc501540849)

[4.3.2 Mapreduce(分布式计算框架) 36](#_Toc501540850)

[4.3.3 Zookeeper(分布式协作服务) 37](#_Toc501540851)

[4.3.4 Pig(基于Hadoop的数据流系统) 39](#_Toc501540852)

[4.3.5 Flume(日志收集工具) 39](#_Toc501540853)

[5 绩效智能决策 41](#_Toc501540854)

[5.1 绩效管理整体结构设计 41](#_Toc501540855)

[5.1.1 KPI主题设计 41](#_Toc501540856)

[5.1.2 绩效管理结构设计 41](#_Toc501540857)

[5.2 基础指标详细设计 42](#_Toc501540858)

[5.2.1 成本 43](#_Toc501540859)

[5.2.2 质量 43](#_Toc501540860)

[5.2.3 设备 43](#_Toc501540861)

[5.2.4 安全 43](#_Toc501540862)

[5.2.5 销售 43](#_Toc501540863)

[5.3 绩效评定流程 44](#_Toc501540864)

[5.3.1 现场需求指标折算 44](#_Toc501540865)

[5.3.2 主题综合评估 45](#_Toc501540866)

[5.3.3 绩效评定 45](#_Toc501540867)

[5.4 KPI分解 46](#_Toc501540868)

[5.4.1 分解流程 46](#_Toc501540869)

[5.4.2 多目标优化算法 47](#_Toc501540870)

[5.5 展示 48](#_Toc501540871)

[5.5.1 展示界面设计 48](#_Toc501540872)

[5.5.2 时序数据展示 48](#_Toc501540873)

[5.5.3 多维层次数据展示 49](#_Toc501540874)

[6 成本智能决策 52](#_Toc501540875)

[6.1 数据基础 52](#_Toc501540876)

[6.1.1 数据源 52](#_Toc501540877)

[6.1.2 各工序成本指标 52](#_Toc501540878)

[6.1.3 数据预处理 54](#_Toc501540879)

[6.2 成本控制与优化 56](#_Toc501540880)

[6.2.1 历史数据追溯分析 56](#_Toc501540881)

[6.2.2 成本指标在线监控 58](#_Toc501540882)

[6.2.3 后工序预警 58](#_Toc501540883)

[6.2.4 成本主题分析整体架构 59](#_Toc501540884)

[6.3 可视化展示 60](#_Toc501540885)

[6.3.1 总界面设计 60](#_Toc501540886)

[6.3.2 图表展示形式 61](#_Toc501540887)

[7 安全主题 63](#_Toc501540888)

[7.1 安全主题数据准备 63](#_Toc501540889)

[7.1.1 数据源确定 63](#_Toc501540890)

[7.1.2 数据清洗技术 63](#_Toc501540891)

[7.2 设备安全分析和预测 64](#_Toc501540892)

[7.2.1 基于安全设施的厂区安全分析 64](#_Toc501540893)

[7.2.2 基于检修率的事故发生预测 65](#_Toc501540894)

[7.2.3 基于综合安全指标的厂区警戒划分 65](#_Toc501540895)

[7.3 人员安全分析和预测 65](#_Toc501540896)

[7.3.1 基于人力资源系统数据的安全分析 65](#_Toc501540897)

[7.3.2 安全反馈表单智能筛选系统 66](#_Toc501540898)

[7.3.3 基于反馈系统数据的安全分析 67](#_Toc501540899)

[7.3.4 结合GPS的员工综合安全评价 67](#_Toc501540900)

[7.4 安全主题评分技术 67](#_Toc501540901)

[7.5 安全主题可视化展示 68](#_Toc501540902)

[7.5.1 设备安全信息可视化 68](#_Toc501540903)

[7.5.2 人员安全信息可视化 69](#_Toc501540904)

[8 销售主题 71](#_Toc501540905)

[8.1 原料产品价格的展示与预测 72](#_Toc501540906)

[8.1.1 原材料价格采购数据的获取及外部数据的爬取 72](#_Toc501540907)

[8.1.2 原材料价格数据的预测模型 74](#_Toc501540908)

[8.1.3 原料价格预测展示界面 77](#_Toc501540909)

[8.2 多主题销售情况分析 77](#_Toc501540910)

[8.2.1 销售数据的预处理 78](#_Toc501540911)

[8.2.2 销售数据的空间分析 79](#_Toc501540912)

[8.2.3 销售数据的时间分析 80](#_Toc501540913)

[8.2.4 客户分析 83](#_Toc501540914)

[8.2.5 界面效果展示 85](#_Toc501540915)

[8.3 产品销量预测 88](#_Toc501540916)

[8.3.1 产品销售数据的预处理 89](#_Toc501540917)

[8.3.2 产品短期销量预测 90](#_Toc501540918)

[8.3.3 界面效果展示 91](#_Toc501540919)

# 项目介绍

高端铝合金功能材料智能制造新模式围绕中铝瑞闽高端铝合金功能材料的产品研发、生产制造、质量管控、仓储物流、优化决策等产品全生命周期的主要过程，建设以“集成化、精益化、数字化、互联化、智能化”为特征的高端铝合金功能材料的智能制造新模式，缩短新材料的研发周期、提高生产效率、提升产品质量、降低过程能耗。通过智能制造新模式的建设，提升行业智能化制造水平，增强企业核心竞争力，实现精益、高效、绿色、智慧生产。

实践表明，通过传统的技术与管理提升，已经难以有效解决制铝企业产品设计、生产制造、经营管理等多个生产与管理环节的全局协调优化问题，也难以解决产品质量、节能减排与生产效益的动态协调与管理控制存在的问题；更加难于满足制铝企业增效、降耗与转型升级的需求。先进的信息技术恰恰可以帮助解决铝企业特有的连续流程整体优化问题，具体包括提升生产效益、保障产品质量、节能降耗、提升生产计划兑现率、减缓机器设备损坏、缩短生产停车时间、减少大修次数等。

随着以社会化网络、移动互联网、云计算等信息技术的兴起和快速发展，数据已经成为社会化的战略资源。一个企业应有数据的规模和运用数据的能力正在成为综合实力和创新能力的重要组成部分。云计算、大数据技术，人工智能在促进工业化与信息化融合方面的重要作用和潜力已显现出来，已成为带动工业和社会发展的重要力量，已经成为驱动铝企业形成创新发展机制，突破增长极限，保障经济快速发展的主要动力。有效地组织和使用大数据将对企业发展与企业创新能力提升产生巨大的推动作用。近年来，云计算、大数据技术日趋成熟，大数据对企业生产、管理已经产生了重大影响；利用大数据技术，进一步提高产品质量及生产效率，降低生产成本及能源消耗，减少排放，实现绿色制造已成为铝企业发展的重要技术措施。大数据是一整套数据分析处理技术体系，更是一种复杂问题解决的思想方法；利用企业生产、经营管理业务数据、产品服务数据、宏观经济数据等构成的“企业大数据”，在企业经营管理决策中开始发挥重要作用。大数据对于企业管理模式创新具有重要影响，是改善企业生产管理能力、提升决策能力、形成企业管理创新的关键。

目前，中铝瑞闽生产包括熔铸、热轧、冷轧、退火及精整五大工序，生产过程涉及多个工序，每个工序均会产生大量的过程数据，包括合同订单信息、产品规范、工艺参数、生产消耗、实绩曲线等。对于这些数据的处理，目前多采用孤立的方式，即仅对单个产品的单个工序进行分析，且数据存储方式简单，保存期间短，没有有效利用这些数据，对生产过程的企业实现精细管理提供有力支持，主要表现在：

（1）不能在产品生产的各工序之间和工序内部实现质量信息的及时传递，出现质量异议，缺乏有效的过程溯源数据，难以准确定位出现问题的环节及快速找到出现问题的原因，导致批量的质量事故或长时间的生产停滞；

（2）由于产品生产过程非常复杂，多种因素耦合在一起，上游工序的生产结果会对下游工序产生遗传影响，简单的数据处理方式及数学模型无法满足高效、高精度的控制要求，也无法为工艺模型的优化提供有效支撑；

（3）产品的营销信息与生产过程数据没有有效关联和融合，不能为企业经营的科学决策提供支持，包括生产组织、资源分配等。

铝合金产品生命周期包括整个生产过程，涉及多个工序，每个工序均会产生大量的过程数据，涵盖铝合金产品生产过程的实时生产数据、产品信息与积累的经验知识，数据具有典型的大数据的“4V+1C”的特征（即Volume，Velocity，Variety，Value和Complexity），其突出的是生产数据以实时流数据为主体，数据量随时间持续快速增长，数据体内部蕴含复杂非线性关系，多源分布异构数据并存等特点；使数据的分析挖掘和应用的难度非常大。生产数据本身的多样性和复杂性及其所表征的铝合金产品生产流程的复杂性，使得企业大数据智能分析与决策技术研究具有典型意义。

因此，本项目研究中铝瑞闽生产制造数据、经营管理数据、营销数据等大数据的多元异构集成、可靠存储、可视化决策分析等关键技术；研究开发基于大数据的决策分析模型、算法，构建模型驱动的企业大数据智能分析与决策支撑平台。面对高端铝合金功能材料定制化、多样化需求，以及制造过程产品一次合格率低、质量异常频繁等问题，项目建设中将在目前企业ERP、APS、MES与PCS等信息与自动化系统架构基础上，通过对制造过程工业大数据的深度利用构建面向产业链内外协同的智能制造决策新模式，构建适应高端定制与灵活交货、多工序协同的智能制造新模式，实现高端产品大规模个性化定制生产，有效降低制造成本，提升产品的价值链和精益服务能力，提升企业对于客户个性化需求的快速响应能力和核心竞争力。

# 项目总体设计

## 设计概述

针对中铝瑞闽生产数据拥有庞大的规模，不断变化的类型，不断演化的分析模式，存在信息量大，信息存储的格式繁多，信息较分散不集中，现有技术难以为铝制产品生产过程进行相应支持，难以对高端产品大规模个性化生产，多工序无法协同智能产生等具体问题，建立中铝瑞闽智能决策系统。

在智能分析与决策技术的基础理论、技术方法、模型算法与应用支撑平台等多个层面，构造综合集成的智能分析与决策技术体系，形成符合制铝生产大数据特点的数据整合与数据分析挖掘技术和智能分析与决策技术平台，建设横跨质量、成本、客户、设备和安全五个主题，基于自定义绩效计算模型的绩效KPI体系。

由于工业过程的复杂性及特点，将智能分析、决策应用与制铝生产系统控制理论相结合，围绕生产流程和工艺模型，研究模型驱动的适用于复杂工业系统特点的大数据整合（如何体现智能、流程、可视化等）、大数据融合、大数据分析挖掘技术。

（1）数据挖掘与分析方法库

针对多对象、多变量、多路径、多证据、多目标的复杂优化问题，提出基于证据的动态规划方法；分别对成本、质量、客户、设备、安全、绩效等KPI指数进行在线监控与分析、及时发现管控中存在的问题、预测变化趋势、分析存在的风险、建立风险预警机制，最终实现成本精益控制、质量精益管控、客户精益服务、设备精益监控、安全精益监督、绩效精益管理。

（2）智能决策业务模型设计

围绕产品全生命周期的核心信息与数据，建立PLM系统分析平台，构建顶层决策支持系统，结合统计分析，分类、聚类，回归预测，多目标优化等智能算法，针对生产调度、成本控制、质量管控、客户服务、设备监控、安全监督、绩效管理等主题业务，分别实现数据主题仓库构建，模型算法设计优化，实现产品全要素、全价值链、全流程、全生命周期的数据与信息集成。

（3）智能决策展示平台

构建整个高端铝功能材料制造流程的KPI导向图，实现数据虚拟分析与实际业务系统的互动，利用数据KPI导向图分析引擎及各数据分析支撑模型，将制造过程各类数据可视化、数字化和智能化，便于决策者和管理者对整个制造过程进行实时掌控，对存在的问题进行及时、准确的决策，提高企业的经济效益。

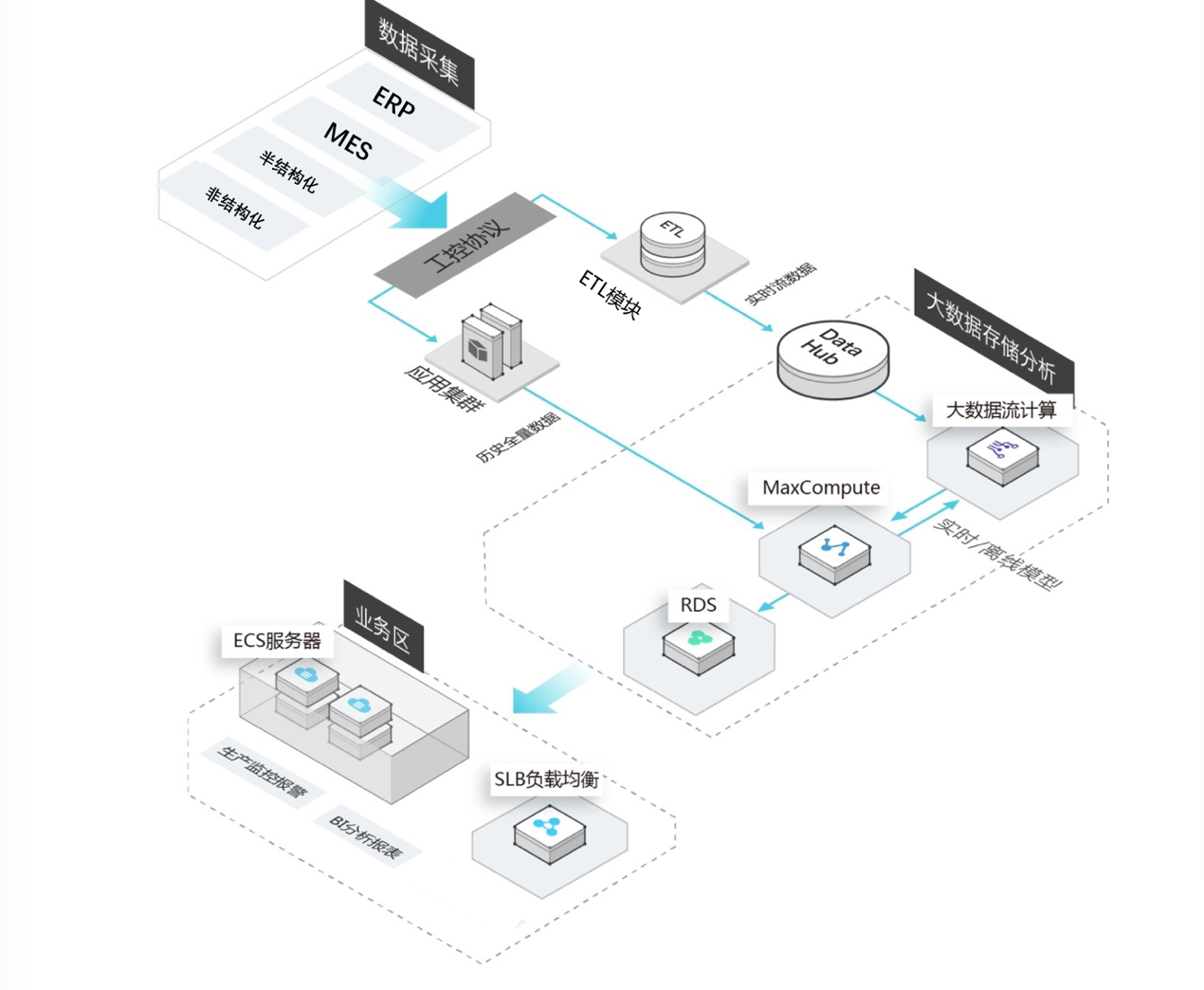


图 ‑1 瑞闽智能决策支持平台

平台逻辑功能框架自底向上是分布式高维数据存储层、元数据服务层、复杂工业模型层、处理分析层以及制铝企业大数据分析处理应用层；同时，需要对分布式系统进行作业、资源调度、管理的协调与监控中间件的支持，支持工作流及其调度的设施。逻辑架构在逻辑上划分为实时数据流处理子系统与大数据智能分析与决策子系统。

实时数据流处理子系统接受实时传感器数据流，数据流元组记录随时间变化的空间信息，具有动态、海量、高维、时效、连续、多源、无限等特性。能够为分析调度、设计规划、异常预警等工业信息管理和决策提供支持，为工业提供更为全面和便捷的决策服务。

大数据智能分析子系统运用适于制铝领域内先进的数据处理技术，根据各工业子系统的需求和它们之间的内在联系，对来自多来源渠道、格式不一致的数据在综合生产信息的基础上进行抽取、集成，并进行深度智能分析与处理，获得可用于决策的模式、模型、规则和知识。

## 项目目标

在智能分析与决策技术的基础理论、技术方法、模型算法与应用支撑平台等多个层面，构造综合集成的智能分析与决策技术体系，形成符合制铝生产大数据特点的数据整合与数据分析挖掘技术和智能分析与决策技术平台，建设横跨质量、成本、设备、安全、客户五个主题，各主题指标量化集成、对接现场需求，各主题综合评定、全厂总体评价四个层级绩效KPI体系，材料工艺参数设计，智能生产高级排程的智能监控决策支持系统。

绩效KPI为不同需求的客户提供适应需求的决策支持服务。包括质量、成本、设备、安全、客户五个主题的指标量化集成，是对各个主题的各项KPI的基本情况展示，包含整个智能决策系统的几乎所有的底层支持信息。

决策支持系统同时包含材料工艺参数设计，智能生产高级排程。材料工艺参数设计采用高端新材料研发新模式，为加快高端铝合金功能材料的研发速度，降低研发成本，提高现有产品的质量稳定性提供支持。智能生产高级排程以客户信息识别为基础，通过客户需求的分解，形成具体订单，并关联物料订单链，基于工艺路径约束和设备能力约束，通过高级计划模型与算法，实现批量生产计划的制定和优化。

### 成本精益控制

成本精益控制包括对熔铸工序、热轧工序、冷轧工序等各工序的核心成本KPI指标进行分析。核心成本KPI指标包括熔炼时的各类物料消耗偏离度和波动率信息，热轧时的耗材消耗、燃料消耗、热轧板成材率及冷轧时的耗材消耗、产品废品率等。

进行基础数据整合、相关数据搜集，数据预处理功能用以对原始数据进行处理，并重整，建立相应的数据之间关系，并按存储模型要求进行存储。

根据目标KPI进行历史数据分析，包括数理统计分析、成本问题挖掘、问题原因追溯等。基于历史数据，统计成本KPI指标统计期望、标准差等特征参数，分析当前指标状态，发现现存成本问题，利用Pearson相关性计算方法，对工序输入、控制、输出参数相关性进行分析，综合使用特征选择、多元回归分析等方法实现问题原因追溯。

### 质量精益管控

质量精益管控包括对熔铸工序、热轧工序、冷轧工序等各工序的核心质量KPI指标进行分析。核心质量KPI指标包括熔铸时熔液成分偏离度、波动率以及熔炼温度，热轧时轧制温度和热轧板质量，冷轧时产品质量等。

实施基础数据整合、相关数据搜集，数据预处理与数据存储、统一数据访问接口。数据采集自熔铸至成品的整个制造流程关键工艺装备、系统中的工艺参数、控制设定参数和物料、质量参数。

根据目标KPI进行历史数据分析，包括数理统计分析、控制问题挖掘、问题原因追溯等。通过对各产线的产品和过程实施过程质量信息采集，统计分析各指标当前状态，与历史数据进行对比，挖掘质量控制问题，基于Pearson相关性计算方法，分析工序输入、控制、输出参数相关性，综合使用特征选择、多元回归分析等方法实现问题原因追溯。

### 客户精益服务

客户精益服务分别包含产品导向、时间导向、客户导向的销售分析。

进行基础数据整合、数据预处理等步骤用以对原始数据进行处理，根据各个分析主题将数据按照不同关键字索引进行归类划分，建立相应的数据之间关系，并按存储模型要求进行存储。

采用不同模型对不同类型的数据进行分析，利用均值、同比变化率、加权得分等方式实现对产品、效益、客户的定量分析，并且针对数据的不同特点，使用折线图、柱状图等方式实现可视化展示，兼具实用性与美观性。对客户进行了评级与打分，让销售部门可以根据不同类型的客户群体提供不同的服务，做到客户的精益化服务。

### 设备精益监控

设备精益监控包括设备性能维护和发挥类指标计算和基于指标的厂区警戒划分。

针对铝生产工序中的多种设备，收集设备维护和生产的设备基本信息，使用领域相关公式完成设备完好率、设备利用率和维修费用率的计算，从多个角度反应设备管理工作的情况。还通过设备的故障信息完成设备工序状况系数计算，依据工序状况系数应用聚类方法实现对厂区设备不同警戒程度的划分。

### 安全精益监督

安全精益监督主要包含员工安全资历计算、员工反馈指数计算和安全反馈表单智能筛选系统。

员工安全资历计算主要基于人力资源系统记录的员工档案数据，通过无监督聚类算法对员工进行聚类，并根据其与中心点的偏移量化得到安全资历指数，将其作为员工的基本信息进行存储。员工反馈指数主要基于安全反馈系统收集的统计记录，通过公式计算得出有效反馈率再结合相应的系数，最终得到员工的有效反馈程度的评价指标。安全反馈表单筛选系统通过关键词模糊分类，基于安全资历和反馈指数的优先级划分以及审核人员的重要性评价完成对整体表单的筛选处理，降低审核人员的工作量。

# 系统框架

## 系统框架总体结构

### 总体逻辑框架

本项目是在中铝瑞闽生产流程范围内，面向生产制造和经营管理相关的分析评价、预测与决策等数据分析挖掘需求，以产品质量管理为主线，针对生产工序间质量分析、生产数据挖掘为目标的节能减排、营销决策支持等具体需求，深度融合实时生产数据、生产管理信息以及工艺知识，研发基于云计算与大数据技术的数据分析挖掘、深度学习、决策分析模型，利用已有成果和开源技术研发模型驱动的可视化、系统化的中铝瑞闽智能分析与决策应用支撑平台。

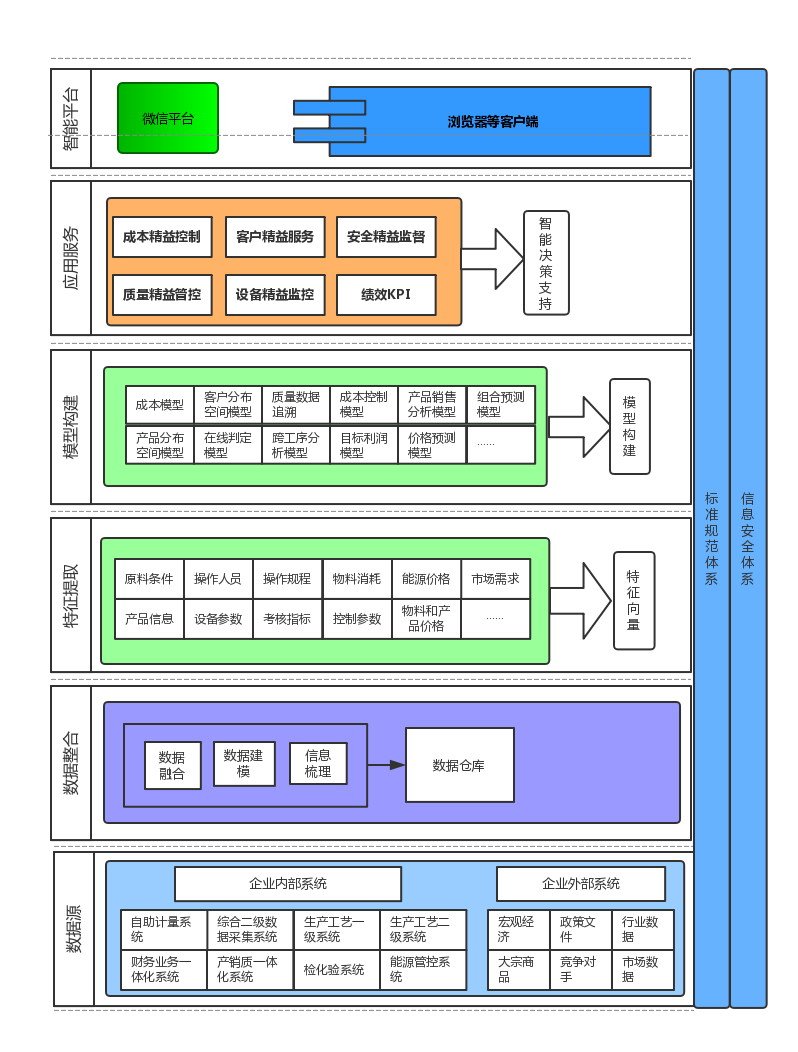


图 3‑1 瑞闽大数据智能决策平台逻辑框架

### 总体技术框架

在本项目中，数据来源可以分为产品数据、运营数据、价值链数据、外部数据这四个方面，不仅数据规模较大，且来源不一。所以最终采用了Hadoop的HDFS和spark的并行计算框架来作为数据处理、存储和分析的平台，构建了包含了数据源、数据整合、模式识别、模型构建、应用服务、智能平台共六个层次的智能决策系统。下面由下至上依次介绍每个层次的实施技术方案。

1. 数据源

数据源包括企业内部系统和企业外部系统，对于企业内部系统，进行实时备份镜像，主题数据仓库从备份镜像数据库中抽取数据；外部数据包括宏观经济、行业数据、市场数据等，采用网络爬虫从多个网络开放数据源进行爬取。

1. 数据整合

数据整合的核心步骤就是构建数据仓库，为后面的知识挖掘和模型构建等数据分析操作提供数据准备。

传统的OLTP（在线事务处理系统）使用了所有的访问数据的操作 CRUD（创建、读取、更新、删除），数据仓库则是完全不同的另一种应用程序，它是用来分析数据并从数据中发现新的价值，它并不是解决所有问题的通用结构，而聚焦于某一类问题，例如质量、成本等。

1. 特征提取

根据质量、成本、客户、设备、安全等主题的KPI，从原始数据或数据仓库中针对问题提取出有效特征用于进行分析，提高分析预测过程中数据处理的效率，增强分析及预测结果的准确性及可靠性。

1. 模型构建

根据不同的业务流程及KPI指标，需要构建相应的模型进行处理。目前行业内广泛使用的数据分析语言是python，而基于python构建的scikit-learn机器学习库提供了大量机器学习算法的实现，能快速实现为对数据的回归、分类、聚类及降维等操作。

1. 应用服务

采用MVC三层结构，构建应用服务系统后台。MVC包括三个层次内容，controller用于控制请求，model利用之前构建的主题模型，根据当前业务需求进行后台服务编程，view层用于客户端显示界面的控制。

应用服务系统需要根据主题及绩效KPI编写成本、质量、客户、安全、设备及绩效KPI六个业务处理模块，模块具有低耦合高内聚的特性。

结合上面提到的需求及实施条件，采用Django这种基于MVC模式的Web框架来构建智能决策系统应用服务后台。Django采用python语言进行开发，可以有效的结合python广泛而有效的社区，及其开发的一系列高效的第三方库资源开展包括数据分析，模型构建，逻辑业务处理等一列列工作。

1. 智能平台

智能决策展示平台使用B/S模式，即服务端和浏览器端，拟采用两种方式进行与决策人员的交互，包括网页浏览器端展示及微信平台的消息推送。

使用浏览器作为访问端增强了系统的灵活性，能够便捷的在不同设备，不同操作系统，不同浏览器设备进行访问，微信平台的信息推送则可以增强信息传递的即时性。

相应的开发任务包括网页开发和移动端开发两部分，使用基于网页的开发语言html、css、javascript，以及借助构建起上的前端开发框架jquery、vue.js等提高开发效率。界面的可视化展示则使用echarts，echarts是网页图表展示组件，提供数据接口，将数据处理后传入相应组件即可获得预期的可视展示效果。

## 数据获取

### 实时镜像备份

采用ADG进行数据库的实时备份，ADG包括的是当前或接近当前的数据，ADG反映的是当前业务条件的状态，ADG的设计与用户或业务的需要是有关联的，ADG的更新是根据业务的需要进行操作的，而没有必要立即更新，因此它需要一种实时或近实时的更新机制。

### 网络爬虫scrapy

对于宏观经济、行业数据、市场数据等外部数据需要从不同的有效网络数据源进行爬取，如何高效统一管理不同类型数据的网络爬虫是需要解决的问题，scrapy具有爬去速度快，爬取功能强大，使用简单的特点，一个典型的scrapy结构如下图所示：



**图3-2 scrapy体系结构**

Scrapy主要包括了以下组件：

引擎(Scrapy): 用来处理整个系统的数据流处理, 触发事务(框架核心)

调度器(Scheduler): 用来接受引擎发过来的请求, 压入队列中, 并在引擎再次请求的时候返回. 可以想像成一个URL（抓取网页的网址或者说是链接）的优先队列, 由它来决定下一个要抓取的网址是什么, 同时去除重复的网址

下载器(Downloader): 用于下载网页内容, 并将网页内容返回给蜘蛛(Scrapy下载器是建立在twisted这个高效的异步模型上的)

爬虫(Spiders): 爬虫是主要干活的, 用于从特定的网页中提取自己需要的信息, 即所谓的实体(Item)。用户也可以从中提取出链接,让Scrapy继续抓取下一个页面

项目管道(Pipeline): 负责处理爬虫从网页中抽取的实体，主要的功能是持久化实体、验证实体的有效性、清除不需要的信息。当页面被爬虫解析后，将被发送到项目管道，并经过几个特定的次序处理数据。

下载器中间件(Downloader Middlewares): 位于Scrapy引擎和下载器之间的框架，主要是处理Scrapy引擎与下载器之间的请求及响应。

爬虫中间件(Spider Middlewares): 介于Scrapy引擎和爬虫之间的框架，主要工作是处理蜘蛛的响应输入和请求输出。

调度中间件(Scheduler Middewares): 介于Scrapy引擎和调度之间的中间件，从Scrapy引擎发送到调度的请求和响应。

Scrapy运行流程大概如下：首先，引擎从调度器中取出一个链接(URL)用于接下来的抓取；引擎把URL封装成一个请求(Request)传给下载器，下载器把资源下载下来，并封装成应答包(Response)；然后，爬虫解析Response；若是解析出实体（Item）,则交给实体管道进行进一步的处理；若是解析出的是链接（URL）,则把URL交给Scheduler等待抓取。

## 数据仓库

### 数据仓库设计

数据仓库建设主要按照组织结构、工艺流程等分大类集中管理分布式存放，其中每一大类都包括实时数据、准实时数据、归档数据等；在此基础上构建共享数据库、发布数据库等主题数据库，各类数据按照变化特点，利用时间、触发条件等进行增量式自动维护管理。

由于数据的使用目的不同、数据的使用频率不同、数据使用方式不同，采用分区管理的方式来存储和管理数据。根据具体应用需求，建立应用数据资源池，主题数据按照应用单独进行数据的划分。主题数据建立采用交互式、定制化的方式建立。可按照主题、事件、对象等建立主题数据。

### 数据仓库管理

数据仓库管理功能包括常规ETL、数据状况监管、数据仓库调度规则管理、数据质量管理等。

数据的管理还支撑数据的使用，主要包括常规的数据探索（各类统计分析、各类图表、常规分析挖掘）、基于流程或业务主题的数据可视化展示、通过规范接口支持数据的建模以及模型库的使用。由于中铝瑞闽智能决策系统数据仓库总量规模较大，数据之间的关系复杂，所以要实现数据的ETL、解决数据清理、实现数据的整体动态增量自动化管理、自动化维护，建立良好的易用可操作数据管理系统是关键。

建立数据仓库的主要任务是建立中铝瑞闽企业全局性、基础性的数据中心。本研究以云平台、Hadoop为技术依托，进行了中铝瑞闽基础数据接入、数据清洗算法、统计口径整理、数据编码中铝瑞闽指标体系设计及实现，最终使中铝瑞闽数据达到指标规范、口径一致、数据字典标准。中铝瑞闽智能决策系统中心数据仓库设计成三个层次，以实现数据的稳定性与应用的灵活性相统一。

研究为中铝瑞闽企业提供面向中铝瑞闽资源开发业务对象，支持数据采集、数据存储、分析处理和决策支持的数据管理体系结构和方法。通过中铝瑞闽智能决策系统数据仓库的建设，实现中铝瑞闽企业数据的统一集中管理、数据共享和数据交流，为中铝瑞闽企业提供数据分析应用环境，使中铝瑞闽管理者全面、及时、准确地掌握企业信息，为中铝瑞闽生产经营管理提供可靠依据，实现生产经营决策的科学性和及时性； 促进和规范中铝瑞闽企业乃至整个铁矿行业的信息标准化建设。

### 数据接入

数据采集是根据中铝瑞闽生产经营决策分析主题，采集操作环境中的各类业务数据，重整归类后放到数据仓库，再经多层次分类汇总成为有效的管理信息。数据采集过程跨越操作环境和信息分析环境，是数据仓库建设中的重要一环。

在中铝瑞闽智能决策系统智能化建设中，对现场数据及业务数据库数据的数据接入提出了更高的要求。按照中铝瑞闽信息化系统的典型设计和公司公司数据接入规范的要求，基于模块化架构，利用Webservice等技术实现了平台通用化数据接入方法，并结合中铝瑞闽工业设备状态监测应用对课题方法进行了改进，可以有效解决中铝瑞闽智能决策系统数据接入方式的软项目需求，为实时/历史数据新平台提供了统一规范化的数据接入，从而提升了结构化、非结构化、实时流数据及业务数据接入与管理的效率。

（1）中铝瑞闽数据特点

中铝瑞闽是一个复杂的动态时空巨系统，其地理空间要素、资源环境信息和生产经营信息的内容广泛、综合、复杂、变化迅速，这就决定了中铝瑞闽数据具有如下特点：

海量数据：由于中铝瑞闽的生产经营管理，涉及产品数据：设计、建模、工艺、加工、测试、维护数据、产品结构、相关的配置关系、变更记录、客户需求、尺寸、标识等；运营数据：组织结构、业务管理、生产设备、市场影响、质量控制、生产、采购、库存、目标计划、电子商务等；价值链数据：客户、供应商、合作伙伴、价格体系、运输管理等；外部数据：经济运行数据、行业数据、市场数据、竞争对手数据等，其数据必然是海量的。

数据的多性与多源性：中铝瑞闽数据来源广泛，既有内部生产运营数据，又有外部相关数据，既涉及技术，又涉及经济等等;数据表现形式有的是以数据文件形式存储，有的是以文本形式存储等。这就要求在进行中铝瑞闽信息系统集成时对数据进行规范和统一。

中铝瑞闽数据的动态性：中铝瑞闽数据对中铝瑞闽地质体的表达是一个由模糊到精确的过程，是一个动态的积累过程，中铝瑞闽数据量越来越大，反映中铝瑞闽实体的层次逐步细化，反映地质体客观现象、规律的准确程度逐步提高。

（2）分布式数据接入

利用数据采集系统，对生产调度、计划统计、财务等系统的上报数据进行接收，存放到原始数据库中。在中铝瑞闽智能决策系统分布式系统环境下，本研究基于Flume建立数据仓库的整个系统的数据收集服务。Flume将有以下主要功能：

1）可扩展性：Flume采用常用的三层软件体系结构，由Agent，Collector、Storage构成。对于其中每层都可进行水平扩展扩容。其中，Master将管理维护全部Agent与Collector，这样将使得整个系统更易于监控与维护，并且可通过使用ZooKeeper来管理与处理负载均衡问题，这也有效的避免了单点故障问题。

2）可管理性：所有Agent、Colletor节点进程均由Master统一管理，这使得系统便于维护。在多个Master情况，Flume将使用ZooKeeper与Gossip协议，从而保证系统动态配置全局一致性问题。使用者也可在Master节点上监控调取各数据源节点或者数据流节点实时工作情况，并可对各数据源配置和动态加载。与此同时，Flume也提供Web 界面与Shell命令行的两种方式对系统及数据流配置进行管理、操作。

3）可靠性：可靠性是分布式系统中需要保障的问题。当系统节点发生不可处理故障时，日志系统需要能被发送到其它正常工作的节点上，以避免发生数据丢失情况。Flume系统为此提供三种级别的配置方式，由强到弱的层次保障数据可靠性。其级别依次分为：端到端方式，当收到数据后Agent节点先将事件写到磁盘上，当数据发送成功后，再将其删除；若数据传送失败，可进行重新发送；Store on Failure方式，若接收数据的节点发生故障时，将数据缓存到本地，待接收数据节点恢复后，恢复并继续发送；Best Effort方式，当数据成功发送到接收方节点后，其不会进行确认操作。

Agent运行时会根据设定创建一个或多个虚拟机，每个虚拟机做为一个进程执行具体的数据收集任务。各Agent数据收集节点相对独立部署，基于Flume的分布式数据收集服务系统架构模型如下图所示：

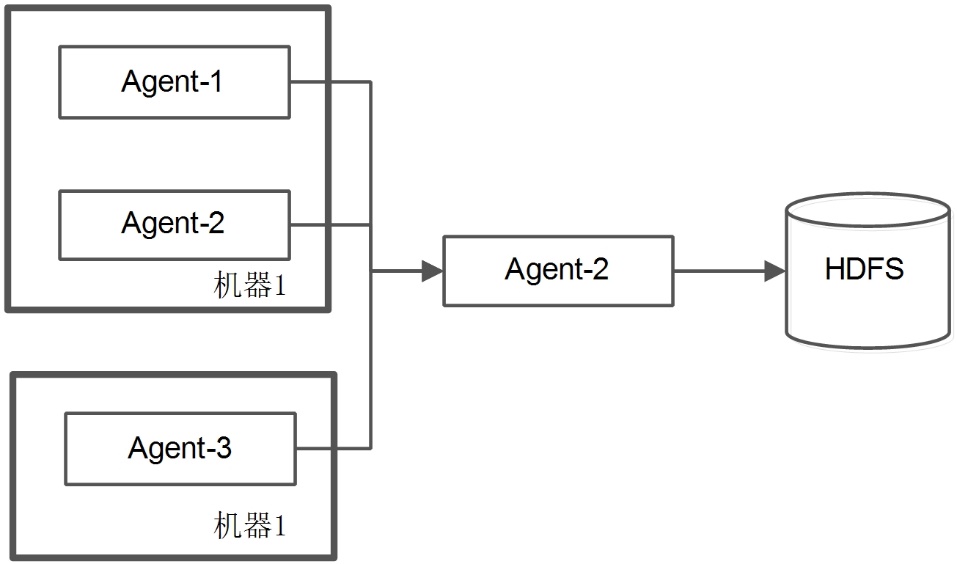


图 3‑3基于Flume的分布式数据收集服务系统架构

针对中铝瑞闽特有的PI实时数据库，采集模块开发了基于C#的PI数据库测点提取系统。此系统可根据测点、时间等维度对实时数据库进行快速提取，并通过RPC将数据接入到Flume系统中。

为支持中铝瑞闽智能决策系统系统各类业务，本采集系统中可对接多种数据源及数据存储类型，其中源数据类型包含：Avro协议、Thrift协议、JMS消息系统、Netcat源、Mysql数据库、Oracle数据库、MSSQL数据库、WebService、文件系统等。采集系统对存储数据类型支持有：分布式文件系统HDFS、IRC、Thrift协议、HBase、Mysql数据库、Webservice等。

### 数据清洗

建立中铝瑞闽智能决策系统的数据仓库目的主要是为联机分析和决策支持等相关应用业务服务，是为了优化企业资源调配、提升企业的协同和资源共享，从而提高企业的竞争力和创新力。高质量的数据能够促进和提高企业决策的正确性，而质量差的数据则起到反面的作用。数据质量的控制成为数据仓库建设发展过程中越来越引起重视的突出问题之一。

与此同时，由于生产环境中的各种原因，导致中铝瑞闽数据系统中，现有数据库存在这样或那样的数据质量问题，主要表现为：不正确的属性值、重复记录、拼写问题、不合法值、空值、不一致值、缩写词不同、不遵循完整性规则、时间序列流数据缺失问题等。此外，从多数据源中抽取数据时，由于各数据源的数据库表结构的设计可能不相同，当完成从多数据源到数据仓库的数据迁移后，同样会产生一些错误或冗余等信息。上述这些问题构成了所谓的数据质量，故需要进行系统的数据清洗。要对中铝瑞闽智能决策系统建设数据仓库时所面临的复杂数据清洗问题，课题提出公司数据清洗系统框架。

本研究的数据清洗框架主要由四模块组成：数据检测、数据清洗、数据评估、元数据管理，其主要特征体现在扩展性和交互性。系统不仅提供了现成多种检测算法、清洗策略及评估方法，而且随条件变化在各个模块中可分别加入自定义的方法和策略，这体现了系统的可扩展性。另外由于数据多样性和复杂性，机器很难自动完成数据清洗过程，这要求人能随时参与其中，这体现了系统的交互性。



图 3‑4数据清洗框架

清洗系统中的各个模块可以批量执行，也可以即时执行。在即时执行过程中，单个检测算法或单个清洗策略的执行效果可以立即呈现给用户，由用户决定是否需要修改检测算法和调整清洗策略，这样做法可以提高准确度。批量执行可以提高效率，但比较难把握准确度，即时执行和批量执行要根据实际情况取得平衡。具体执行流程如图所示。

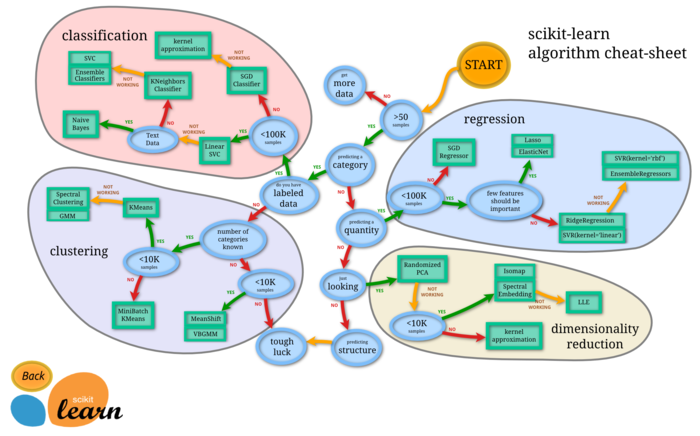


**图 3‑5数据清洗执行步骤**

## 知识挖掘和模型构建

python的scikit-learn内置丰富的算法和分析模型，同时结合google开源的tensorflow，可以快速高效的实现神经网络算法。

scikit包含丰富的模型，包括贝叶斯、随机森林算法、线性鉴别分析（LDA）、稀疏回归、逻辑回归、稀疏线性、特征建模、决策树分裂、邻接点分析等，模型概览及选择可参考下图：



**图3-6 scikit-learn算法参考**

同时可以使用tensorflow构建神经网络模型，一个完整的tenflow构建图计算如图所示：



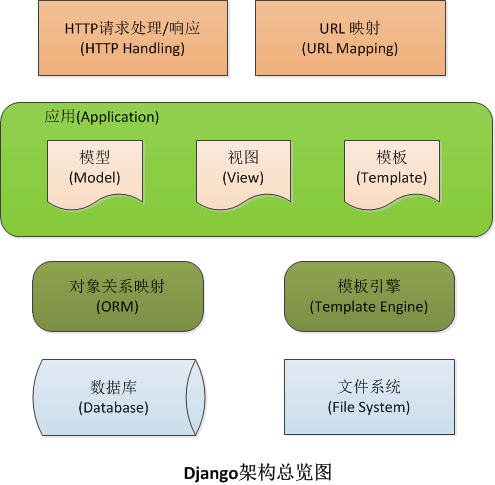
**图3-7 tensorflow图计算**

## 应用服务与智能平台

### 应用服务后台

智能决策应用服务后台由MVC模式的Web框架Django实现。

Django 是一个高级的 Python 网络框架，可以快速开发安全和可维护的网站。由经验丰富的开发者构建，因此可以专注于编写应用程序，而无需重新开发。Django是免费开源的，有一个繁荣昌盛而积极的社区，在应用程序开发过程中遇到的问题都可以去社区中讨论以进行解决。Django总体结构如下图所示：



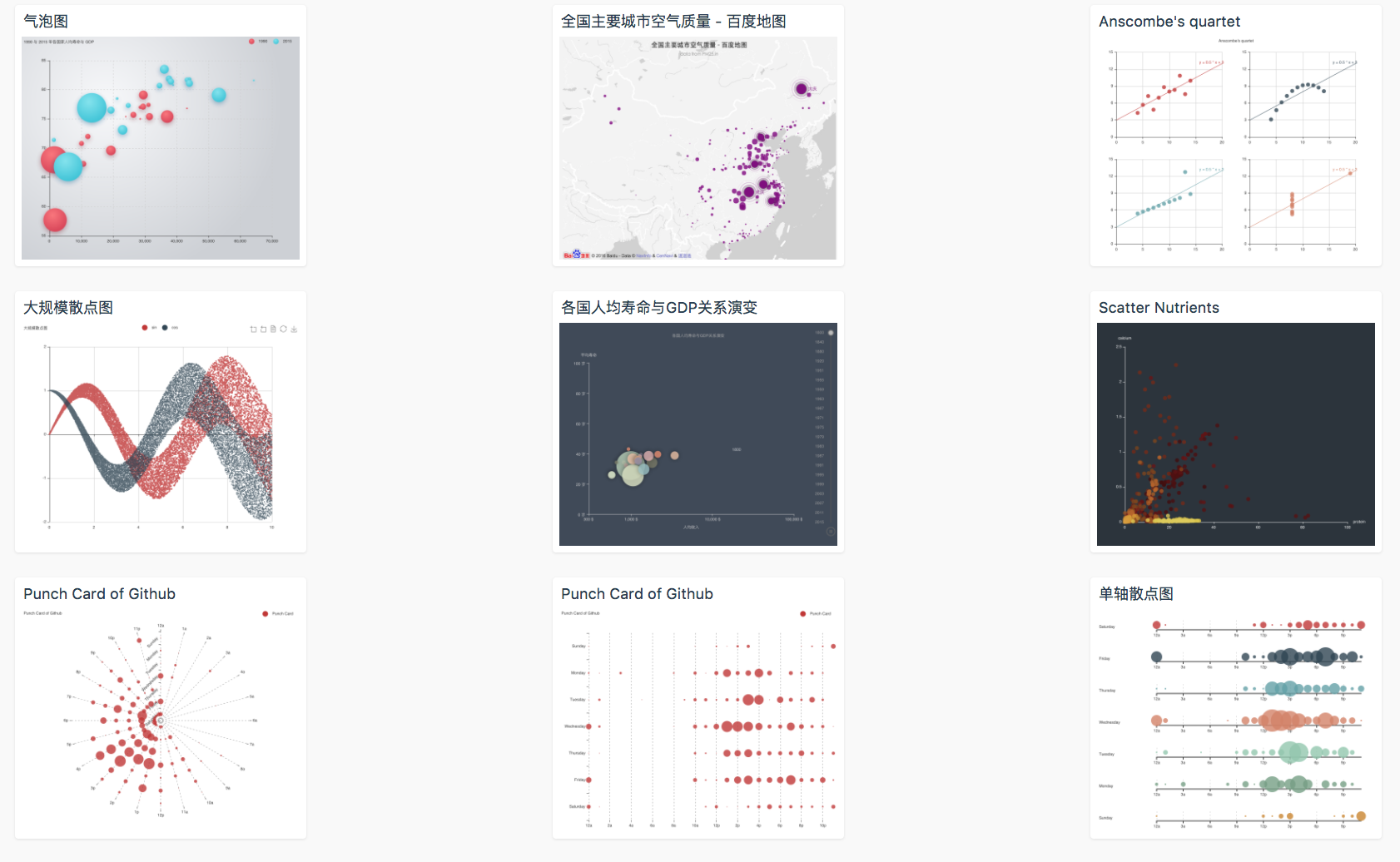
**图3-8 Django结构概览**

### 智能可视化平台

平台可视化展示使用网页形式展现，图表展示使用echarts组件库。echarts是一款开源免费的web组件，它的功能丰富，涵盖各行业图表，满足各种需求，同时它也具有活跃的社区，在应用开发过程中遇到的问题都可以得到有效的解答。

ECharts 提供了常规的折线图，柱状图，散点图，饼图，K线图，用于统计的盒形图，用于地理数据可视化的地图，热力图，线图，用于关系数据可视化的关系图，treemap，多维数据可视化的平行坐标，还有用于 BI 的漏斗图，仪表盘，并且支持图与图之间的混搭。

可以在下载界面下载包含所有图表的构建文件，如果只是需要其中一两个图表，又嫌包含所有图表的构建文件太大，也可以在在线构建中选择需要的图表类型后自定义构建。常用的echarts图表如下图所示：



**图3-9 echarts常见图表展示**

## 数据服务平台

从“数据+服务”体系的角度出发，中铝瑞闽智能决策支持系统作为一个大型智能决策支持系统，仅仅关注数据的生产是远远不够的，必须从“使用者”的需要、习惯、喜好出发，以“持续满足使用者对决策支持的需求、期望”为目标，通过建立各种信息服务手段，把智能决策支持系统所能提供的数据“包装”成满足使用者需要、贴合他们习惯的“服务”，并通过多种渠道提供给使用者，才能真正体现中铝瑞闽智能决策支持系统的价值。这就是中铝瑞闽智能决策支持系统的以服务为核心的出发点。

在设计以服务为核心的业务时，首先要对其业务特征进行分析，然后才能有的放矢的进行业务过程的设计。

首先，用户对“服务”的需求是个性化、多变的，无止境的，缺少衡量标准的。每个用户都是一个独立的个体，有着与别人不同的服务预期和使用习惯，不同类别不同个体的用户对服务的方式和内容有不同的要求，即便是同一用户在不同环境或不同时间对服务很可能有不同的要求和期待，因此其需求是个性化、多变的。随着用户的使用和中铝瑞闽数据资源库的发展，使用者对中铝瑞闽智能决策支持系统的服务预期会不断提高，中铝瑞闽智能决策支持系统的服务对象也会更加广泛且可能是日益增加的，因此对“服务”需求是无止境的，这显然对中铝瑞闽智能决策支持系统的服务提供机制的扩展能力提出很高要求，要能以最小成本应对无止境的需求增长。此外，对于服务效果、服务水平“好坏高低”的直接反映，很大程度上是由使用者的主观感受（即服务体验）所决定的，而非单纯的技术，因此缺少硬性的衡量标准，需要在服务过程中不断的去探查、摸索，主动的收集使用者对服务的需求（虽然过程很困难），不断的从用户反馈中去总结，从而主动的改进服务以使用户不断的提升服务体验。

其次，“服务”强调主动出击、快速反应、全程管理。一个好的服务提供机制从来不是一个单纯“被动响应”的过程，而是一个“主动反应”的过程，主动的收集用户对服务的需求以进行相关服务的设计与提供来实现主动的服务能力扩展，主动的在服务提供的全过程中去收集用户的使用情况信息，快速的调整服务策略和服务提供形式，并根据用户的行为，分析用户的潜在需要，主动把相关数据、相关服务推送/推荐给用户。这样，用户在使用服务的过程中会不断的在系统的主动引导下发现自己所需（包括潜在需要）的服务能力，不断地深入挖掘系统的服务能力，从而带动整个中铝瑞闽智能决策支持系统服务水平的提升。

最后，“服务”与“数据”是紧密相关的。一方面数据是服务的基础（可以把“数据”看成是“服务”的原料，没有数据，服务也就成了“无源之水”），另一方面“服务”业务也扩展了“数据”业务，使数据业务从单纯“数据层面”的数据产品生产升级为“面向服务需要”的数据资源生产，从而扩展了数据业务各环节的业务内涵。此外，用户对于“服务”的需求归根结底都会反映到“数据”上，因此对于“服务”的需求也会转化为对“数据”的需求，会引发数据“生产”过程的变化和扩展。

因此，在上述业务特征基础上的中铝瑞闽智能决策支持系统“以服务为核心”的业务就是一个包含服务设计与提供，全程主动管理并与数据紧密联系的过程，如下图所示：



图 3‑10 以服务为核心”的业务流程图

在上图所示的流程中，首先是服务的规划设计，通过服务的规划设计，对服务过程进行管理。一方面收集服务需求，进行对应的服务设计，并通过信息服务与推送环节提供给使用者，实现服务手段的扩展和个性化。另一方面将服务需求中对数据的要求转化为对数据的需求，对“数据”生产过程“提要求”，从而“改变/扩展”数据生产过程，使之满足服务的需要。这样，整个业务环节就形成一个可自增长的“闭环”（这也是“服务”不断进化的体现），各业务环节的业务领域是相对稳定的，但业务内涵和细节随着服务流程的循环在不断变化、扩展（即全动态的业务）。

通过这样的“服务需求的收集-分析-设计-提供-反馈”的循环过程，不断的扩展、丰富信息服务手段，就可以使中铝瑞闽智能决策支持系统具备最长久的生命力，不断去满足用户个性化、无止境的需要。

## 系统信息安全

目前，公司认识到了信息安全的重要性，并建立了信息安全防护体系。然而，由于基础薄弱，工作人员防范意识淡薄，信息安全形势不容乐观，计算机网络系统瘫痪、数据丢失、网上信息失窃等重大信息安全事故时有发生，给企业造成不必要的损失。一般来说，国际的一般作法是信息安全的投入要占到企业信息建设投入的15％。对于不同的行业，比例会有所不同。随着信息技术的飞速发展，信息安全的含义从保障系统的稳定运行发展到全面促进业务开展。互联网的发展已使中铝瑞闽行业的业务开展突破了时间和空间的限制，大大推进了业务的发展，甚至在一定程度上已改变了业务模式。而所有新技术的运用都伴随着信息安全风险的产生。随着信息技术进一步深入业务层面，信息安全与业务开展的关系也越来越紧密。

在互联网条件下实现数据安全交互与安全管理是一项复杂的系统工程。北京科技大学知识工程北京市重点实验室与北京易恒信公司合作研发的双因子组合公钥（TF-CPK）密码体制已经得到国家密码管理局审查批准；在北京市科委的支持下开发出具有世界领先水平的基于双因子组合公钥（TF-CPK）的标识认证系统。目前，TF-CPK标识认证技术已经发展成熟，所开发的产品已经有成熟的应用，为建设基于CPK的企业级信息安全体系提供了坚实的技术基础。

CPK标识认证系统的基础是我国自主创新的组合公钥（CPK）密码体制。2007年5月，经严格审查，该密码体制在西班牙巴塞罗召开的欧洲密码年会上首次对外介绍，得到国际承认。2008年11月，双因子组合公钥（TF-CPK）密码体制正式通过国家密码管理局的审查批准。基于标识的认证技术也列入美国2009年度联邦政府科研计划。而CPK技术则率先实现解决了这一难题，其认证规模理论上可达到10的48次方以上，并支持端到端基于标识的直接认证，认证过程不需要管理中心支持因而不受中心管理能力、设备水平以及数据库、带宽等外部条件限制。

CPK提供了标识和种子密钥这两种重要的管理资源。CPK标识认证系统是直接对标识进行证明的系统；当用户端和服务端进行相互认证时，服务端马上就能确认对方的身份和权限，并据此提供服务或进行限制。因此，只要建立起科学的标识管理体系，设计部署好“基因”，系统就能自动实现安全管理。通过对标识进行科学定义，保证其唯一性，即可构建与现行管理体系相适应，便于直接识别、应用的认证体系。这是CPK标识认证系统独有的特性，其对构建大规模的复杂体系十分关键。

基于CPK标识认证技术体系通过将数据的使用控制管理及对数据加密两种手段紧密结合，构建起一个从服务器到客户端，涵盖存储、传递到使用各个环节，集数据加密、访问控制、授权管理、动态密钥交换、过程审计、动态跟踪为一体的信息安全支撑体系。不仅能够能保证数据在存储状态下和网络传输过程中的安全，还可以通过数字签名保证数据的真实性、完整性和不可抵赖性，实现数据密码级的安全存储、传递和共享。

### 基于CPK数字签名技术的身份认证方案

我们有了完善的数字签名、验证签名的方案，就可以在以上方案的基础上设计出基于CPK的随机数的签名体系的具体的登陆认证方案。

方案的具体数据流程主要是：

1）用户通过用户名和密码登录系统；

2）Server接受到请求后产生随机数通过用户对应的标识进行加密回送到用户浏览器；

3）用户浏览器通过activeX控件调用解密接口解密随机数；

4）然后用密钥设备（CPK Key或CPK Card）对当前会话的随机码进行数字签名并发送至服务器；

5）服务器接收到用户提交的数字签名，调用CPK认证API进行验证，如果用户是该作用域允许的合法用户，则可以通过验证；

6）数字签名验证通过后可以进一步对签名的时间戳，最后从数字签名中取得用户的登录信息，交由原系统做进一步的业务逻辑处理。

### 基于CPK数字签名技术的登陆记录不可抵赖方案

系统在数据库字段上实现了用户登陆的不可抵赖性，在每次进行登陆验证通过之后，都可以把CPK算法的签名数据、服务器产生的随机数、用户标识和操作记录绑定在数据库中，而通过CPK算法的验证签名过程，我们可以知道，系统在提取每一项登陆历史记录的时候，都可以用CPK算法验证一次用户标识和签名数据，如果验证成功，说明确实是该用户登陆了移动远程控制系统，该用户无法抵赖这次登陆行为。

### 基于CPK的文件加密解密方案

首先需要更改数据库存储文件的方式，原来的明文存储改成密文，用户文件信息表中的需要多加一个字段RandomSeed（它表示随机数种子），由系统自己随机生成，生成后立即使用加密算法对文件进行加密，将密文文件存于服务器对应位置，随机数种子由用户标识进行加密并存储于服务器对应的文件表信息上，系统开发者和数据托管中心都无法知道随机数是什。

在服务器解密方面，用户申请解密读数据时，系统将随机数种子的明文发送到客户浏览器，由客户浏览器通过控件和用户输入的pin码启动CPK设备，解密随机数种子回送到服务器，服务器利用解密的种子对文件进行解密，显示到用户浏览器，至此完成数据的服务器解密。

### 基于CPK的安全数据库方案

以CPK标识认证系统为基础，将访问控制手段与数据库加密技术有机结合，形成涵盖存储、传递到使用各个环节的严密的伦理审查信息安全共享体系，保证即使出现极端的情况（如数据库服务器被搬走），也不会直接导致数据库内容的泄露。

实现严格的过程控制和管理，以CPK标识认证系统为基础，将访问控制、授权管理、过程审计、动态跟踪等纳等功能纳入系统设计，形成一体化的管理体系。实现跨域交叉认证，通过对种子密钥的管理实现不同信任域和管理域之间的用户实现端到端（包括服务端与客户端，客户端与客户端之间）的跨域交叉认证，解决不同信任域之间的认证登录。全过程加密，利用CPK标识认证系统基于标识的密钥交换能力，实现数据在存储、上传及下载过程中的加全密状态，从根本上保证数据库的安全。保证服务端效率，利用CPK密钥交换算法和数字信封技术解决服务器的动态密钥管理分配问题，避免服务器对数据进行频繁的加解密操作，保证系统的高效运行。实现端到端的安全数据存储和交互，利用CPK基于标识的密钥交换能力，通过客户端应用软件的开发，建立标准化的基于标识的密钥交换机制，实现客户端数据的安全存储（利用自己的标识进行加密），和客户端与服务端、客户端与客户端数据的安全交互（利用对方的标识进行加密），从而实现数据库共享的私密化（利用某用户标识生成的公钥加密只有用该用户的标识私钥才能解读）。

## 消息推送服务平台

智能决策支持微信平台是基于腾讯公司微信企业号、RTX即时通讯、企业邮箱和腾讯云等产品，以连接一切为核心，统一以智能决策支持结果发布与互动为主要内容提供通讯录、即时通讯、邮箱、校内通知、交流互动、生产状况上报、决策支持结果推送、信息资源等多项管理和应用为一体的智能决策支持私有云服务平台。

智能决策支持微信平台为公司用户提供移动应用入口。平台连接单位与个体间相互认可且稳定的关系，通过完全匹配的组织架构和单位通讯录，可以与现有IT应用一致的唯一用户账号，为企业用户提供统一移动应用入口。

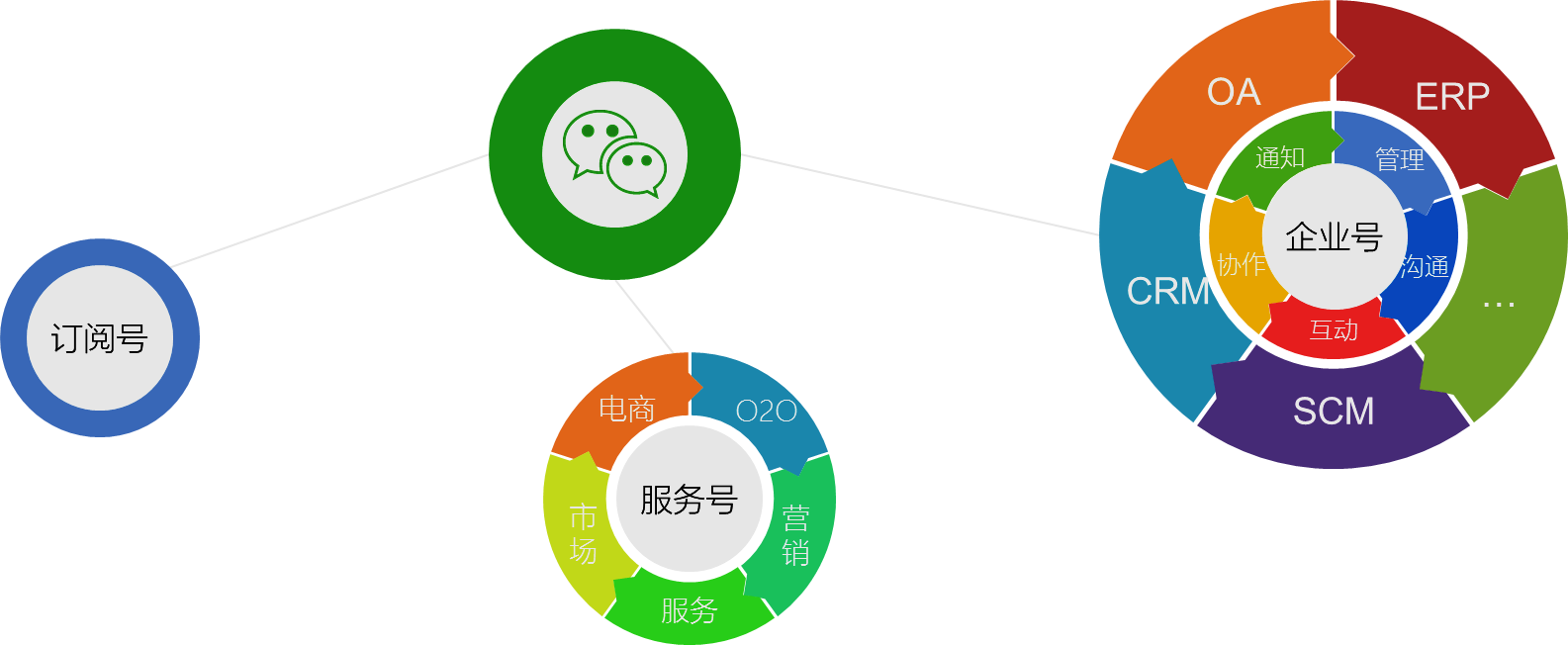


图 3‑11 微信平台组织架构

鞍钢公司微信平台定位为“公众信息发布的平台、专业信息查询的窗口”，在发布公司相关信息的同时，作为公司职工的专业用户，可实现生产信息查询、网络问企互动、安全隐患上报等各应用系统的信息查询及会议互动等功能。

微信公众平台平时不光要随时查看用户留言，适时回复之外，最主要的就是发布生产信息和生产决策信息、发布生产信息和生产决策判定的实时消息，通过这个平台对公司公司进行多方面的宣传和推广，让更多的人群关注企业生产状况。而生产信息和生产决策信息的选择、形式、版式、格式等都应严格规范。

# 大数据与系统集成

## 工业大数据平台

针对中铝瑞闽在生产过程及企业内部管理过程中产生的海量、多源、异构数据，提出基于Hadoop的HDFS为主文件系统，关系型数据Mysql、Oracle与文档型数据库混合存储为基础存储层的工业大数据分析平台，如下图。

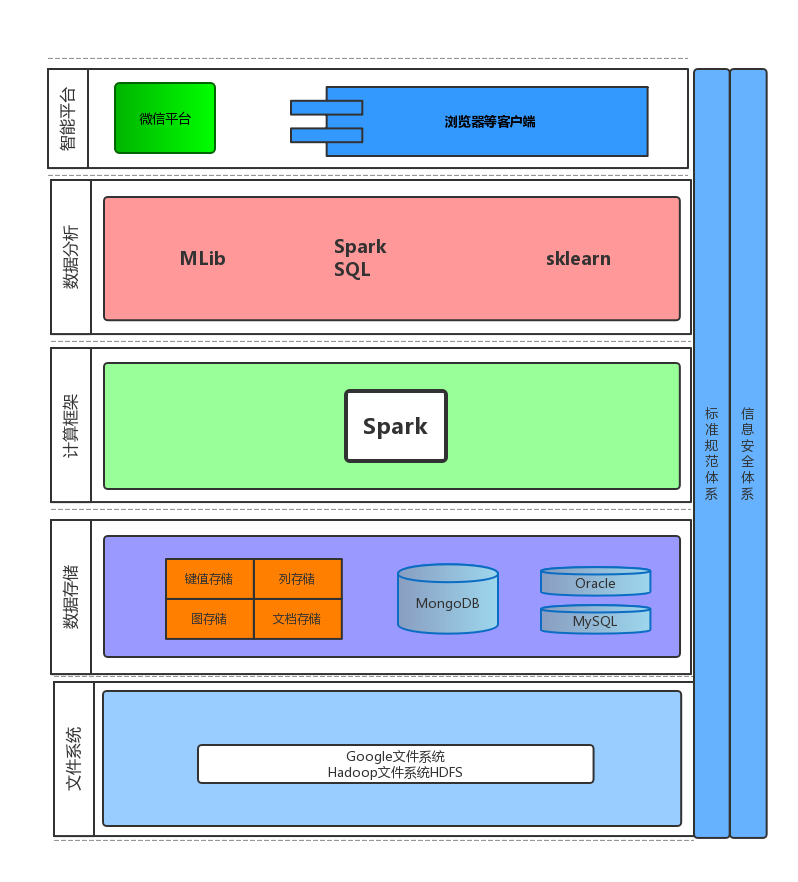


图 4‑1 工业大数据解决方案

## Spark分布式计算框架

### Spark概述

spark生态圈以spark为核心，以弹性分布式数据集（RDD）为基础，打造了一个基于内存计算的大数据平台。伯克利将Spark的整个生态系统称为伯克利数据分析栈（BDAS）（见图2）。

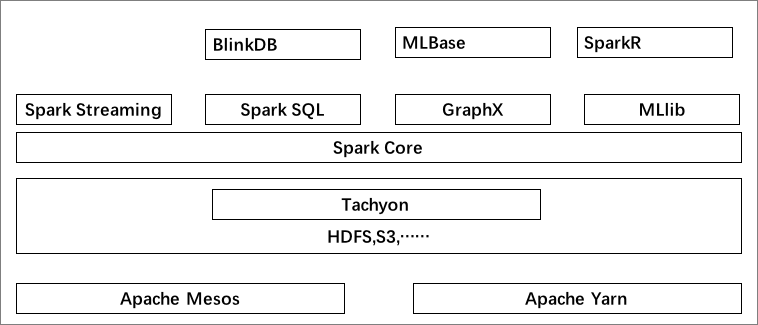


图4‑2 BDAS结构图

与其他大数据技术，特别的与hadoop的MapReduce相比，spark具有以下优势：

提供了一个全面、统一的框架用于管理各种有着不同性质（文本、图表数据等）的数据集和数据源（批量数据或实时的流数据）的大数据处理需求；

内存数据共享。

Apache Spark是一种包含流处理能力的下一代批处理框架，与Hadoop的MapReduce引擎基于各种相同原则开发而来的spark主要侧重通过完善的内存计算和处理优化机制加快批处理工作负载的运行速度。

Apache Spark本身没有一个分布式存储系统，因此spark需要一个第三方的分布式存储，因而需要为其配备相应的分布式存储环境，由于hadoop已经在行业内有着多年的应用，在各方面有着完备的生态，使用HDFS文件系统存储数据，可用于任何兼容于hadoop的数据源，包括HDFS，Hbase，Cassandra。所以选取hadoop的HDFS作为文件系统，即将spark与hadoop集成，保留HDFS的同时使用spark取代MapReduce引擎进行计算。Spark将中间的计算结果保存在内存中，具有更高的并行执行效率并且更适合于数据挖掘和机器学习等需要迭代处理的领域，因而契合了即时或是类即时查询的需求。

### MLlib数据挖掘算法库

MLlib是Spark项目子模块，提供一些可扩展的机器学习领域经典算法的实现，旨在帮助开发人员更加方便快捷地创建智能应用程序。MLlib包含许多实现，包括聚类、分类、推荐过滤、频繁子项挖掘。此外，通过使用PySpark库，MLlib 可以有效地扩展到云中。

MLlib的主要目标是创建一些可扩展的机器学习领域经典算法的实现，旨在帮助开发人员更加方便快捷地创建智能应用程序。MLlib现在已经包含了聚类、分类、推荐引擎（协同过滤）和频繁集挖掘等广泛使用的数据挖掘方法。除了算法，MLlib还包含数据的输入/输出工具、与其他存储系统（如数据库、MongoDB 或Cassandra）集成等数据挖掘支持架构。

MLlib提供一些可扩展的机器学习领域经典算法的实现，旨在帮助开发人员更加方便快捷地创建智能应用程序。MLlib 包含许多实现，包括集群、分类、推荐过滤、频繁子项挖掘。

## Hadoop生态系统

### HDFS(Hadoop分布式文件系统)

HDFS是一个分布式文件系统。整个HDFS系统设计了两套自己的协议，都是基于TCP/IP协议之上设计的：Client Protocol和DataNode Protocol。Client Protocol负责用户端与文件系统的通信，而文件系统内部各个节点之间通过DataNode Protocol协议来实现内部的通信和文件和管理。HDFS采用了主从(Master/Slave)结构模型，一个HDFS由一个NameNode和若干个DataNode组成，其中NameNode作为主服务器，管理文件系统的命名空间和用户端的连接。集群中的DataNode则管理各自存储的数据。HDFS 内部的所有通信都基于标准的TCP/IP 协议。

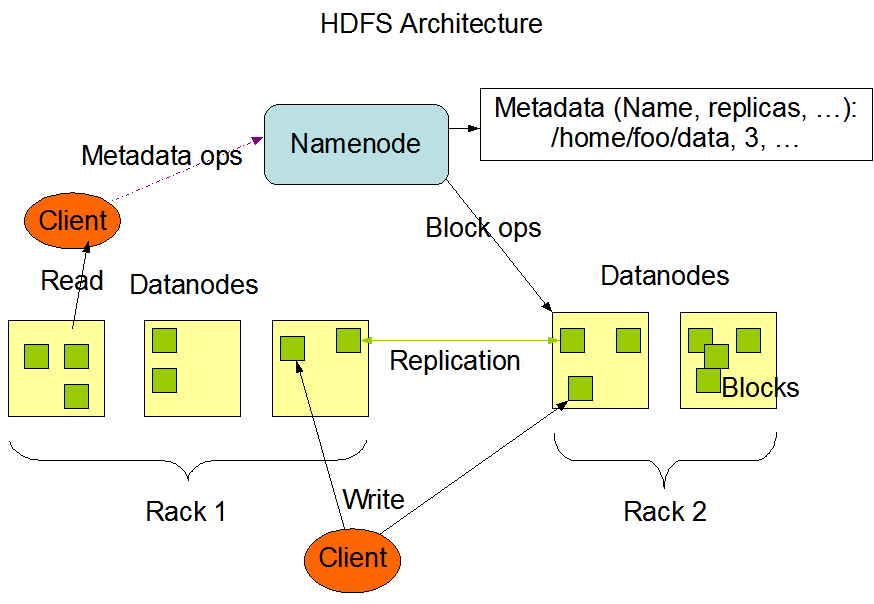


图 4‑3 HDFS结构图

**Client：**切分文件；访问HDFS；与NameNode交互，获取文件位置信息；与DataNode交互，读取和写入数据。

**NameNode：**Master节点，在hadoop1.X中只有一个，管理HDFS的名称空间和数据块映射信息，配置副本策略，处理客户端请求。

**DataNode：**Slave节点，存储实际的数据，汇报存储信息给NameNode。

**Secondary NameNode：**辅助NameNode，分担其工作量；定期合并fsimage和fsedits，推送给NameNode；紧急情况下，可辅助恢复NameNode，但Secondary NameNode并非NameNode的热备。

### Mapreduce(分布式计算框架)

MapReduce是一种编程模型，用于海量数据集的并行运算，其架构如图5所示。概念“Map（映射）”和“Reduce（化简）”，是它的主要思想，都是从函数式编程语言里借来的，还有从矢量编程语言里借来的特性。它极大地方便了编程人员在不会分布式并行编程的情况下，将自己的程序运行在分布式系统上。MapReduce在执行时先制定一个map函数，把输入键值对映射成新的键值对，经过一定处理之后交给reduce，之后reduce对想同key下所有的value进行处理再输出键值对作为最终的结果。其框架实现是由一个单独运行在主节点上的JobTracker和运行在每个集群从属节点上的TaskTracker共同组成。主节点负责调度构成一个个作业，这些作业分布运行在从属节点上，主节点监控它们的执行情况并管理失败的作业重新执行。

MapReduce是一种计算模型，用以进行大数据量的计算。其中Map对数据集上的独立元素进行指定的操作，生成键-值对形式中间结果。Reduce则对中间结果中相同“键”的所有“值”进行规约，以得到最终结果。MapReduce这样的功能划分，非常适合在大量计算机组成的分布式并行环境里进行数据处理。

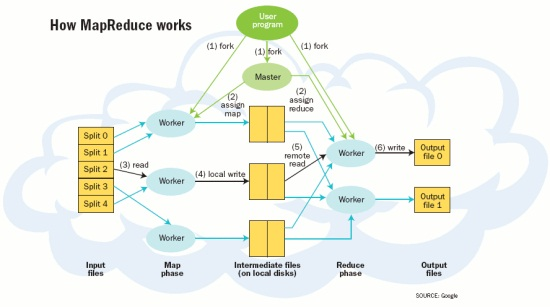


图 4‑5 MapReduce结构图

**JobTracker：**Master节点，只有一个，管理所有作业，作业/任务的监控、错误处理等；将任务分解成一系列任务，并分派给TaskTracker。

**TaskTracker：**Slave节点，运行Map Task和Reduce Task；并与JobTracker交互，汇报任务状态。

**Map Task：**解析每条数据记录，传递给用户编写的map(),并执行，将输出结果写入本地磁盘(如果为map-only作业，直接写入HDFS)。

**Reducer Task：**从Map Task的执行结果中，远程读取输入数据，对数据进行排序，将数据按照分组传递给用户编写的reduce函数执行。

### Zookeeper(分布式协作服务)

源自Google的Chubby论文，发表于2006年11月，Zookeeper是Chubby克隆版。解决分布式环境下的数据管理问题：统一命名，状态同步，集群管理，配置同步等。

ZooKeeper是一个[分布式](http://baike.baidu.com/view/402382.htm)的，开放源码的[分布式应用程序](http://baike.baidu.com/view/553502.htm)协调服务，是[Google](http://baike.baidu.com/view/105.htm)的Chubby一个[开源](http://baike.baidu.com/view/9664.htm)的实现，是Hadoop和Hbase的重要组件。它是一个为分布式应用提供一致性服务的软件，提供的功能包括：配置维护、名字服务、分布式同步、组服务等。

ZooKeeper的目标就是封装好复杂易出错的关键服务，将简单易用的接口和性能高效、功能稳定的系统提供给用户。

ZooKeeper包含一个简单的原语集，提供Java和C的接口。

ZooKeeper代码版本中，提供了分布式独享锁、选举、队列的接口，代码在zookeeper-3.4.3\src\recipes。其中分布锁和队列有Java和C两个版本，选举只有Java版本。

ZooKeeper是以Fast Paxos算法为基础的，paxos算法存在[活锁](http://baike.baidu.com/view/281236.htm)的问题，即当有多个proposer交错提交时，有可能互相排斥导致没有一个proposer能提交成功，而Fast Paxos作了一些优化，通过选举产生一个leader，只有leader才能提交proposer，具体算法可见Fast Paxos。因此，要想弄懂ZooKeeper首先得对Fast Paxos有所了解。

ZooKeeper的基本运转流程：

1、选举Leader。

2、同步数据。

3、选举Leader过程中算法有很多，但要达到的选举标准是一致的。

4、Leader要具有最高的zxid。

5、集群中大多数的机器得到响应并follow选出的Leader。

在Zookeeper中，znode是一个跟Unix文件系统路径相似的节点，可以往这个节点存储或获取数据。如果在创建znode时Flag设置为EPHEMERAL，那么当创建这个znode的节点和Zookeeper失去连接后，这个znode将不再存在在Zookeeper里，Zookeeper使用Watcher察觉事件信息。当客户端接收到事件信息，比如连接超时、节点数据改变、子节点改变，可以调用相应的行为来处理数据。Zookeeper的Wiki页面展示了如何使用Zookeeper来处理事件通知，队列，优先队列，锁，共享锁，可撤销的共享锁，两阶段提交。

那么Zookeeper能作什么事情呢，简单的例子：假设我们有20个[搜索引擎](http://baike.baidu.com/view/1154.htm)的[服务器](http://baike.baidu.com/view/899.htm)(每个负责总索引中的一部分的搜索任务)和一个总服务器(负责向这20个搜索引擎的服务器发出搜索请求并合并结果集)，一个备用的总服务器(负责当总服务器宕机时替换总服务器)，一个web的cgi(向总服务器发出搜索请求)。搜索引擎的服务器中的15个服务器提供搜索服务，5个服务器正在生成索引。这20个搜索引擎的服务器经常要让正在提供搜索服务的服务器停止提供服务开始生成索引，或生成索引的服务器已经把索引生成完成可以搜索提供服务了。使用Zookeeper可以保证总服务器自动感知有多少提供搜索引擎的服务器并向这些服务器发出搜索请求，当总服务器宕机时自动启用备用的总服务器。

### Pig(基于Hadoop的数据流系统)

由Yahoo!开源设计，动机是提供一种基于MapReduce的ad-hoc(计算在query时发生)数据分析工具。

Pig是一种数据流语言和运行环境，用于检索非常大的数据集。为大型数据集的处理提供了一个更高层次的抽象。Pig包括两部分：一是用于描述数据流的语言，称为Pig Latin；二是用于运行Pig Latin程序的执行环境。

Apache Pig 是一个高级过程语言，适合于使用 Hadoop 和 MapReduce 平台来查询大型半结构化数据集。通过允许对分布式数据集进行类似 SQL 的查询，Pig 可以简化 Hadoop 的使用。

用MapReduce进行数据分析。当业务比较复杂的时候，使用MapReduce将会是一个很复杂的事情，比如你需要对数据进行很多预处理或转换，以便能够适应MapReduce的处理模式。另一方面，编写MapReduce程序，发布及运行作业都将是一个比较耗时的事情。Pig的出现很好的弥补了这一不足。Pig能够让你专心于数据及业务本身，而不是纠结于数据的格式转换以及MapReduce程序的编写。本质是上来说，当你使用Pig进行处理时，Pig本身会在后台生成一系列的MapReduce操作来执行任务，但是这个过程对用户来说是透明的。

### Flume(日志收集工具)

Cloudera开源的日志收集系统，具有分布式、高可靠、高容错、易于定制和扩展的特点。

**可靠性。**当节点出现故障时，日志能够被传送到其他节点上而不会丢失。Flume提供了三种级别的可靠性保障，从强到弱依次分别为：end-to-end（收到数据agent首先将event写到磁盘上，当数据传送成功后，再删除；如果数据发送失败，可以重新发送。），Store on failure（这也是scribe采用的策略，当数据接收方crash时，将数据写到本地，待恢复后，继续发送），Best effort（数据发送到接收方后，不会进行确认）。

**可扩展性。**Flume采用了三层架构，分别为agent，collector和storage，每一层均可以水平扩展。其中，所有agent和collector由master统一管理，这使得系统容易监控和维护，且master允许有多个（使用ZooKeeper进行管理和负载均衡），这就避免了单点故障问题。

**可管理性。**所有agent和colletor由master统一管理，这使得系统便于维护。多master情况，Flume利用ZooKeeper和gossip，保证动态配置数据的一致性。用户可以在master上查看各个数据源或者数据流执行情况，且可以对各个数据源配置和动态加载。Flume提供了web和shell script command两种形式对数据流进行管理。

**功能可扩展性。**用户可以根据需要添加自己的agent，collector或者storage。此外，Flume自带了很多组件，包括各种agent（file， syslog等），collector和storage（file，HDFS等）。

它将数据从产生、传输、处理并最终写入目标的路径的过程抽象为数据流，在具体的数据流中，数据源支持在Flume中定制数据发送方，从而支持收集各种不同协议数据。同时，Flume数据流提供对日志数据进行简单处理的能力，如过滤、格式转换等。此外，Flume还具有能够将日志写往各种数据目标（可定制）的能力。总的来说，Flume是一个可扩展、适合复杂环境的海量日志收集系统。

# 绩效智能决策

## 绩效管理整体结构设计

绩效管理则指各级管理者和员工为了达到组织目标共同参与的绩效计划制定、绩效辅导沟通、绩效考核评价、绩效结果应用、绩效目标提升的持续循环过程，其目的是持续提升个人、部门和组织的绩效。

绩效智能决策体系基于人员设备安全层、设备工序运行层、生产过程监控层和客户服务层四个层面，囊括成本、质量、设备、安全、销售五个主题，集成指标信息，评估企业绩效，对公司运行现状进行评价，规划下一步企业发展目标。

从关键业绩指标体系制定、绩效评定和绩效优化三个方面来构建绩效智能决策体系。首先根据公司实际情况制定合理的关键业绩指标体系，建立绩效计算模型，实时获得绩效评分结果以反映当前企业的运行状况是好是坏，当绩效评分偏低时，自动进行多目标优化，提供各指标的优化参考值，以供用户进行过程决策和实时决策。

### KPI主题设计

绩效管理的核心为KPI绩效考核。KPI即关键业绩指标，是通过对组织内部流程的输入端、输出端的关键参数进行设置、取样、计算、分析，衡量流程绩效的一种目标式量化管理指标，是把企业的战略目标分解为可操作的工作目标的工具，是企业绩效管理的基础。

根据瑞闽铝厂的实际情况，将KPI指标分为成本、质量、设备、安全和销售五个主题。

对这五个主题的KPI指标进行监控，建立计算模型，实时进行绩效评定，并以折线图、雷达图、漏斗图等形式展示评定结果。

### 绩效管理结构设计

如图1-1所示，绩效管理模块采用4+1层结构设计实现，分为绩效评定和绩效优化两大部分。

绩效评定由四层构成：

第一层为基础指标，包含代表成本、质量、设备、安全、销售五个主题的基础关键指标。

第二层为现场需求指标，为企业实际制定的KPI，可由基础指标折算而来。

第三层为主题综合评分，包括五个主题各自的评分，每一主题的评分都由相应的现场需求指标计算得来。

第四层为绩效评分，综合各个主题的评分得到最终的绩效评估结果。

针对绩效评定流程，系统单独设计自定义模板子模块，该模块提供增加、删除指标的功能，并定义加、减、乘、除、求和、求平均值、加权、统计最大值、最小值等计算功能模块，使用户可以通过选择现有的指标和计算符号，输入相应参数来自定义计算模型。

KPI分解子模块根据用户指定的下一阶段的绩效目标，使用多目标优化算法，获得各层指标优化值，为用户做决策提供参考。

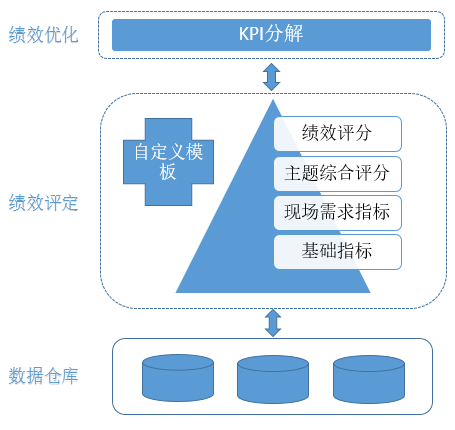


图 5‑1 绩效管理结构图

## 基础指标详细设计

绩效评定流程的关键在于现场需求指标的确定，现场需求指标都从基础指标中衍生而来，基础指标是指能够表征各个主题运行现状的关键参数。因此，基础指标的选取必须具有代表性，同时为了尽可能满足各种现场需求，基础指标必须具有相对完备性，即保证企业提出的各种现场需求指标，都可以通过基础指标计算得到。基础指标大体上划分为成本、质量、设备、安全、销售五个主题部分，每一个指标的数据都在系统人员设备安全、设备工序运行、生产过程监控和客户服务四个层面处理之后的基础上，对数据仓库中的数据进行统计、量化得到。

### 成本

成本主题的基础指标包含原料消耗、合金和渣料消耗、燃料消耗、铝合金溶剂、精炼剂、冷却剂、各工序用水量、耗电量、轧辊刀具等易磨损设备的折旧费用、定期或非定期进行的设备维护、检修费用、成材率折算、设备折旧等成本。

### 质量

质量主题的基础指标包含产品一次检验合格率、原铝合格率、出厂产品合格率等。

### 设备

设备主题的基础指标包含设备总量、完好设备数目、故障停机次数、设备实际开动时间、故障停机时间等。

### 安全

安全主题的基础指标包含安全设施完备评价、设备检修覆盖率、设备健康指标、设备安全分级、安全权重、厂区警戒划分、员工安全资质评价和分级、具体设备最低操作人员安全资质要求、反馈报告评价指数、员工有效反馈率、反馈贡献度评判得分、员工提交等级划分、员工位置安全指数等。

### 销售

销售主题的基础指标包含利润率，退货率，平均帐期，顾客满意度，月销售额，季销售额，季新客户数量、客户投诉次数，比较去年同期的月销量变化，比较去年同期的季销量变化，年销量变化、铝制品市场份额所占百分比等。

## 绩效评定流程

企业的绩效管理是现代企业人力资源的一个重要组成部分，而绩效考核又是绩效管理中一个关键环节，考核体系的优劣直接影响到绩效管理的结果。绩效评定流程由四层组成，从基础指标出发，一层一层的往上计算，依次计算出现场需求指标值和各个主题综合评分结果，最后得到绩效的评估结果。同时以折线图、饼图、雷达图等方式实时展示绩效评定中各项指标和评分结果，为过程决策提供数据基础。

### 现场需求指标折算

现场需求指标是企业为了进行绩效考核而设计的关键业绩指标（KPI），分为成本、质量、设备、安全、销售五个方面。企业绩效的评分直接由这些指标的值演算而来，选取合理的KPI指标，则得到的绩效评分能够很好的体现企业的运营情况，为企业的决策提供正确的引导，反之则可能会对企业决策造成误导。因此，选取合理的KPI指标，是做好绩效管理的关键。KPI确立的是关键绩效指标，而不是目标，因此，KPI的设定并不是越多越好，而需要抓住绩效特征的根本，科学的设定KPI的考核指标。关键绩效指标强调对企业业绩起关键作用的指标，而不是与企业经营管理有关的所有指标。为了更切合企业的实际情况，现场需求指标应使用瑞闽铝厂指定的KPI。

根据瑞闽铝厂的实际情况，初步设计现场需求指标为：

1. 成本
2. CPI等。
3. 质量
4. 过程控制质量、体系运行监控质量等。
5. 设备
6. 重大设备停机时间、设备完好率、计划检修完成率、设备精度控制等。
7. 安全
8. 制度完备率、安全保卫执行率等。
9. 销售

任务目标完成度、利润增长率等。

现场需求指标由基础指标折算得来，具体折算方式视情况而定，以设备完好率为例，它直接由完好设备数目除以设备总数目得到，如式（1）所示。不同的现场需求指标计算方法不同，因此，每一个现场需求指标都应拥有一个对应的计算方法。

(5-1)

考虑到企业在发展的过程中，其现场需求指标可能会发生变动，可通过自定义模板子模块增加、删除指标并自定义输入相应的计算方法。

### 主题综合评估

主题综合评估基于现场需求指标值，对各个主题进行打分，以反映该主题的状态是否良好，评分取值范围从0-100，0为最差，100为最优，数值越高说明状态越好。根据现实情况，指定某一时间间隔（1小时或者1天），按照特定的计算方法，对前一段时间内的KPI数据进行统计计算，分别得到五个主题的综合评分，并使用折线图等方式对评分结果进行时序上的可视化展示，便于用户实时观察企业各个主题部分的变动情况。具体的计算模型采用企业提供的计算方法。

以成本主题为例，假定成本现场需求指标值为，则成本主题的综合评分为：

(5-2)

其中为具体的计算公式，可由自定义模板子模块自定义输入。

### 绩效评定

绩效是指组织、团队或个人，在一定的资源、条件和环境下，完成任务的出色程度，是对目标实现程度以及达成效率的衡量与反馈。绩效的评定结果直接表征了整个企业的运作状态，是判断企业运作状态是良态还是病态的重要依据。整体绩效评分通过综合计算各个主题的评分得到，通常采用的计算方法为加权平均，具体的计算模型应以企业提供的为准。

使用加权平均等计算方法，对各个主题的综合评分进行计算得到最终的绩效评分，评分取值0-100，0为最差，100为最优，采用折线图等方式展示绩效评分的时序数据，以供用户决策。

假定成本、质量、设备、安全、销售五个主题的评分分别为，则绩效评分等于：

(5-3)

其中为具体的计算公式，可由自定义模板子模块自定义输入。

## KPI分解

根据企业制定的KPI指标，系统按照绩效评定流程从底层一步一步结算数据，评估出企业的总体绩效分值，与用户指定的绩效目标进行对比，当绩效未达到指定目标时，自动进行KPI分解流程，区别于绩效评定中一层一层往上汇总计算的过程，KPI分解从绩效出发，自上而下依次计算出下一阶段各层评分和指标的目标值，然后对当前各评分和指标的现状进行分析，指明下一阶段的发展方向，为用户决策提供参考值，实现绩效的智能决策。

### 分解流程

KPI分解过程是指为了获得既定绩效，使用多目标优化算法一层一层的往下求各层评分和指标的优化解，直到得出一组基础指标Pareto解的过程。KPI分解的最终目的就是实现绩效优化，为达到指定的绩效目标提供优化方向。

用户根据当前绩效状况制定下一阶段企业的绩效目标，以该目标为优化目标，采用加权、线性规划、非支配排序多目标遗传、粒子群等算法构建多目标优化模型，首先求解出下一层面各个主题评分的优化解，然后从五个主题出发，分别求解出各个主题现场需求指标的优化解集，最后再根据现场需求指标的优化解，求解出与之相关联的基础指标的最优解集。计算过程中，忽略静态指标，只求取动态指标的优化值。最终KPI分解得出一组各层级动态指标的Pareto解，指明下一阶段各指标的发展方向。

将多目标优化得到的Pareto解与当前各评分和指标数值进行对比，计算差值，对差值从大到小排序，设定阈值，认为超过阈值的差值为主要影响因素，在界面中输出这些主要影响因素，提示用户这些主题和指标可能是需要重点进行整改的部分。

在KPI分解的成果展示上，以虚线的形式延伸各指标原有的时序折线至下一时间节点，下一时间节点的值设为KPI分解得出的优化参考值，在直观上展示各指标下一阶段的参考调整方向，同时说明当前各个主题指标哪些呈现病态，哪些状态良好，为过程决策提供参考。

### 多目标优化算法

目标优化问题一般地就是指通过一定的优化算法获得目标函数的最优化解。当优化的目标函数为一个时称之为单目标优化(Single-objective Optimization Problem, SOP)。当优化的目标函数有两个或两个以上时称为多目标优化(Multi-objective Optimization Problem, MOP)。不同于单目标优化的解为有限解，多目标优化的解通常是一组均衡解。

多目标优化算法归结起来有传统优化算法和智能优化算法两大类。

1. 传统优化算法包括加权法、约束法和线性规划法等，实质上就是将多目标函数转化为单目标函数，通过采用单目标优化的方法达到对多目标函数的求解。
2. 智能优化算法包括进化算法（Evolutionary Algorithm，简称EA）、粒子群算法（Particle Swarm Optimization, PSO）等。

假定当前绩效评分为70，下一阶段目标评分为85，以粒子群算法为例，多目标优化执行流程如下：

1. 初始化粒子群。设置种群规模为m，在搜索空间中随机初始化每个解的成本、质量、设备、安全和销售评分，计算适应函数值，得到粒子的个体极值和和全局最优解；
2. 根据个体极值和全局最优解，更新每个粒子的成本、质量、设备、安全和销售评分；
3. 重新评估粒子的适应度函数值，更新粒子的个体极值和全局最优解；
4. 判断适应函数值是否达到目标评分要求，若达到要求，则输出全局最优解，否则转至步骤2。

## 展示

### 展示界面设计

系统主要从时间和多层次两个层面来对绩效智能决策体系中的指标和评分数据进行数据分析和实时展示。展示界面示意图如图1-5-1所示，数据展示部分包括空间多维数据和时序数据的展示，其中空间多维数据展示主要是对各评分和指标数据的综合展示，以漏斗图、雷达图等形式将5个主题的评分结果集中显现于一个图中，与KPI分解结果进行对比，直观显示当前各个主题的优劣状况，并以树的形式综合体现绩效评定部分各层级之间数据的关联情况。而时序数据展示则主要以折线图、柱状图等形式展示各评分和指标数据随时间的变化状况，同时根据KPI分解结果呈现出下一阶段的目标值，将时序数据展示划分成四栏，对应于绩效评定流程中的四个层级，每一栏显示对应层级中的评分或指标数据。此外，界面中设置绩效目标设定模块供用户对下一阶段目标进行设置和修改，根据用户设置的绩效目标，实时调整图表中KPI分解结果，并在KPI分解结果分析栏中实时显示对多目标优化结果的分析。



图 5‑2 展示截面示意图

### 时序数据展示

采用柱状图、折线图等方式展现各个指标以及各评分结果的时序数据。在时序数据中，以虚线的形式延伸数据至下一时间节点，来呈现下一阶段的发展目标，也就是KPI分解结果。

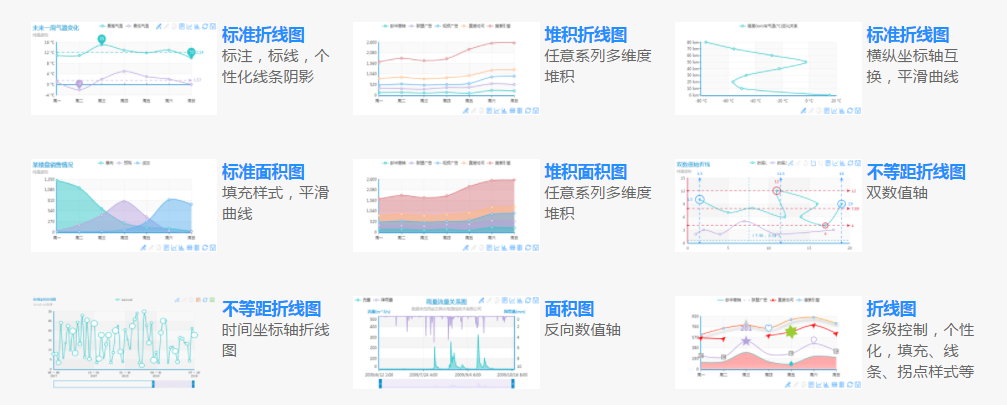


图 5‑3 时序数据展示形式

### 多维层次数据展示

1. 使用饼图、雷达图等形式综合展示成本、质量、设备、安全、销售五大主题评分数据，直观显示各个主题评分对比，揭露出当前企业发展短板所在。可以结合时间轴，绘出虫洞图使用户能一目了然的了解企业各个主题评分的变动情况和好坏趋势。



图 5‑4 雷达图

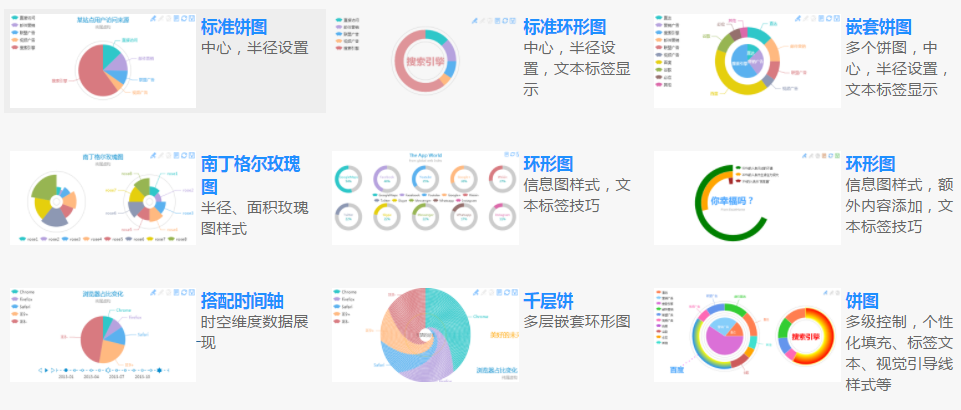


图 5‑5 饼图

1. 结合时间轴线使用折线图、柱状图、散点图、K线图、饼图、雷达图展示多维数据的动态变化。

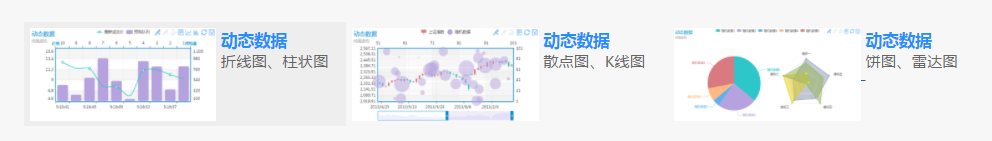


图 5‑6 动态数据展示形式

1. 使用漏斗图、仪表盘等形式展示绩效达标程度，体现当前状态与目标状态的差值，帮助用户快速发现问题，及时调整策略。

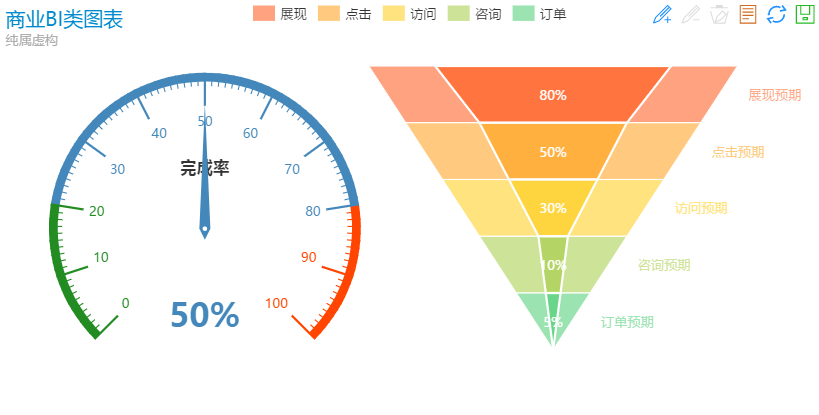


图 5‑7 BI类图表

1. 以树状关系网络形式展现各层次间的数值关系，使用根节点代表绩效，第一层子节点代表5大主题，第二层子节点代表现场需求指标，叶节点代表基础指标，当鼠标移至某一节点时显示该节点信息，包括名称、当前值以及下一阶段目标值。



图 5‑8 层次数据展示形式