中铝瑞闽智能制造新模式应用项目

技术方案规划书

2017-12-29

北京科技大学

目录

[1 项目介绍 2](#_Toc504044428)

[2 项目总体设计 5](#_Toc504044429)

[2.1 设计概述 5](#_Toc504044430)

[2.2 项目目标 6](#_Toc504044435)

[2.2.1 成本精益控制 6](#_Toc504044436)

[2.2.2 质量精益管控 6](#_Toc504044437)

[2.2.3 客户精益服务 7](#_Toc504044438)

[2.2.4 绩效精益管理 7](#_Toc504044439)

[3 系统设计 9](#_Toc504044440)

[3.1 总体架构设计 9](#_Toc504044441)

[3.1.1 系统框架总体设计 9](#_Toc504044442)

[3.1.2 系统软硬件规划 11](#_Toc504044443)

[3.2 数据获取 12](#_Toc504044444)

[3.2.1 实时镜像备份 12](#_Toc504044445)

[3.2.2 网络爬虫scrapy 12](#_Toc504044446)

[3.3 数据仓库 14](#_Toc504044447)

[3.3.1 数据仓库设计 14](#_Toc504044448)

[3.3.2 元数据管理 15](#_Toc504044449)

[3.3.3 数据清洗 17](#_Toc504044450)

[3.3.4 ETL工具 18](#_Toc504044451)

[3.4 知识挖掘和模型构建 18](#_Toc504044452)

[3.5 应用服务与智能平台 19](#_Toc504044453)

[3.5.1 应用服务后台 19](#_Toc504044454)

[3.5.2 权限模块 20](#_Toc504044455)

[3.5.3 智能可视化平台 20](#_Toc504044456)

[3.6 大数据与系统集成 21](#_Toc504044457)

[3.6.1 产品生命周期管理 21](#_Toc504044458)

[3.6.2 大数据计算 22](#_Toc504044459)

# 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，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。

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

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。

## 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，提供成本、质量、服务、绩效四个主题的决策支持，以及基于自定义模版的绩效计算模型。针对高端铝合金制造面临个性化定制和供应链协同需求，以构建的制造过程工业大数据为驱动，通过构建良好的企业与客户之间协同关系，提高企业对客户定制化需求响应速度和管控能力；同时综合利用数据分析和挖掘技术手段等，从制造过程工业大数据中动态获取和深度优化各类工艺规范、约束规则、设备能力和制造成本等知识和规则，以构建虚/实一致的制造流程资源/能力抽象模型，动态反映制造过程资源/能力约束条件和执行情况；决策者利用动态更新的资源来综合分析、评估、分配制造过程的能力，提高决策合理性和客户满意度；管理者通过对制造过程监控，动态响应制造过程各类外部事件，满足客户定制化需求。

# 系统设计

## 总体架构设计

### 系统框架总体设计

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。

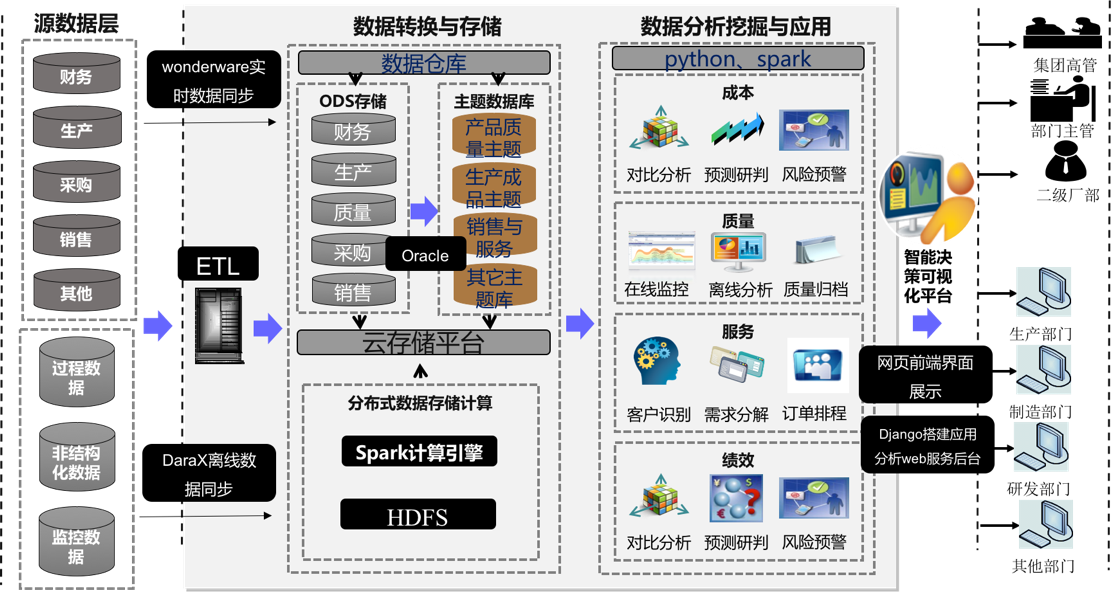


图 3‑1瑞闽智能决策平台架构设计

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Oracle数据库和Hadoop的HDFS构建云存储平台，使用python和spark并行计算框架进行数据处理和分析，构建包含数据源、数据采集交换、数据存储、数据管理、数据应用分析、智能展示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。

平台框架自底向上分为六层，分别是源数据，数据采集转换、数据存储、数据管理和数据分析应用（数据管理层包括元数据管理和主数据管理，分析应用包括模型方法、业务分析和可视展示）和智能展示平台；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。

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. 数据采集交换

将数据从数据源采集到目标数据模型中，通常的操作是ETL（英文 Extract-Transform-Load），用来描述将数据从来源端经过抽取(extract)、转换(transform)、加载(load)至目的端的过程。ETL 是构建数据仓库的重要一环，用户从数据源抽取出所需的数据，经过数据清洗，最终按照预先定义好的数据仓库模型，将数据加载到目标存储位置中去。

1. 数据存储

数据存储ODS、EDW和HDFS由3部分构成。

1. 数据管理

建立统一数据管理门户，对智能决策系统的所有数据进行有效的管理，为瑞闽的管理人员提供便捷的业务、信息与数据的访问渠道。

1. 数据应用分析

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. 智能展示平台

* 后台应用服务

应用服务系统需要根据主题及绩效KPI编写成本、质量、服务及绩效KPI等业务处理模块，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。

* 前端可视化展示

智能决策展示平台使用B/S模式，即服务端和浏览器端，采用网页浏览器进行与决策人员的交互，提供决策支持。

### 系统软硬件规划

表3-1 软硬件需求表

|  |  |
| --- | --- |
| 层级 | 实现方案 |
| 数据管理 | 关系型数据-Oracle、非结构化数据-HDFS |
| 开发语言 | Python3.5 |
| 数据分析 | python工具模块、spark计算集群 |
| 服务后台搭建 | web服务框架Django1.9 |
| 客户端 | 网站页面，echarts(3.0)进行网页内嵌图表展示 |
| 硬件需求 | 1台应用服务器，1台数据库服务器，部分数据存入大数据平台 |

## 数据源

### 企业内部数据

整合企业内部现有信息系统数据源，获取与成本、质量、服务、绩效相关的数据，包括MES、TMS、PCS、二级、OA质量评审数据及APS等。

### 企业外部数据

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

## 数据采集转换

### 实时数据流

实时数据流使用Wonderware进行同步，Wonderware实现了多种业务或工厂应用之间的消息交换和自动数据同步。它是一款商用现成软件，允许业务系统与制造系统之间以一种松散耦合、但高度可靠的方式相互补充。

### 离线历史数据

DataX 是一个异构数据源离线同步工具，致力于实现包括关系型数据库(MySQL、Oracle等)、HDFS、Hive、MaxCompute(原ODPS)、HBase、FTP等各种异构数据源之间稳定高效的数据同步功能。

## 数据存储

数据存储ODS、EDW和HDFS由3部分构成：

* ODS实时同步源系统的数据；
* EDW是企业级数据仓库；
* HDFS是分布式存储平台，存储非结构化和半结构化数据。

### ODS

ODS为源系统的镜像库，且其有较多的并发访问，可以采用传统的关系数据库承担。

可以采用全量备份的形式，使用Oracle的导出工具，全库导出数据进行备份。备份的周期可以根据需要灵活变动，一般是每周进行一次备份，保留一个月左右的备份数据。

可以采用增量备份的形式，导出Oracle的日志增量文件进行备份。通常需要保留近两个月的备份数据。

### EWD

基于决策支持及综合管理系统架构规划设计，整个数据仓库主要规划为五个区域，分别为临时区、主题数据区、应用集市区、数据实验区和数据质量区。

临时区用于暂时存放构建主题库的中间数据；

主题数据区根据瑞闽的业务流程，包括产线各个流程（熔炼、热轧、冷轧等）、产品销售及客户订单等主题，针对主题对所有源系统数据进行整合、归类。

应用集市区是面向应用个性化定制，用于满足不同的分析和管理业务。按照业务发展和应用设计开发的要求，建立各个业务应用数据集市。

数据试验区为数据挖掘过程的数据探索进行数据验证和算法验证，同时满足业务人员或数据科学家提供数据分析过程中临时数据或结果数据保存空间。

数据质量区是用以保存数据质量检查规则和检查结果。数据质量区一般是在数据仓库系统运行较为稳定后（一年后）开始数据质量区的建设，以便保证数据仓库中数据质量的稳定和数据准确。通过建立数据质量检测管理平台，来对数据进行监测和预警。

### HDFS

hadoop主要用于存储非结构化数据、监控数据、历史数据。

## 数据管理

建立数据管理平台，可以实现产品与服务数据管理的一体化和规范化，从而形成统一、规范的数据流，为大数据分析提供数据支撑平台；可以实现产品制造流程的标准化和智能化，从而实现产品全流程制造各环节的高度协同和整体优化；可以实现系统间信息交换的标准化和模块化，从而实现各子系统之间的无缝连接，确保信息交换和传递过程的畅通、准确。

### 报表查询

在数据充足的基础上，可自动定期完成相关的报表统计工作，并在数据管理可视化平台供相应权限的用户查看。

用户也可以根据需求，在交互界面选择需要查看的护具及功能模块，进行实时的报表查询。

### 权限管理

针对中铝瑞闽不同层级的管理人员，在应用服务系统中开发权限系统，基础指标包括成本、质量、服务三个主题的KPI指标，汇总自每个主题的分析结果，面向部门级管理人员，直观展现企业底层业务运行现状；主题综合指标是指针成本、质量、服务三个主题的总体评价性指标，面向集团管理人员，基于系统构建的数据仓库，利用特定的计算方式获得个主题综合指标数值，以供管理人员对企业整体状况进行评估。

## 数据分析

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：

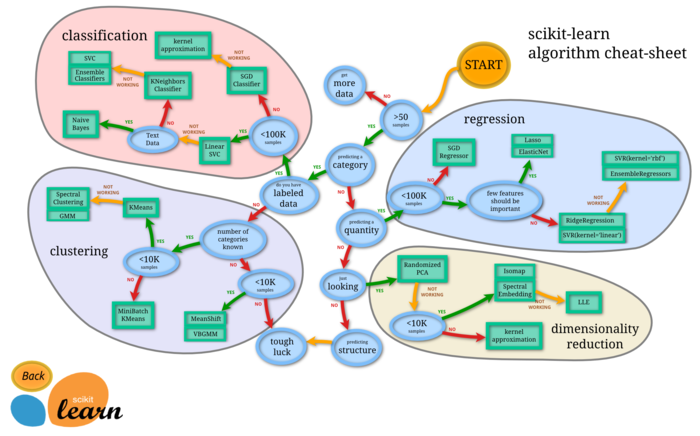


图 3‑3 scikit-learn算法参考

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 3‑4 tensorflow图计算

分析的主要内容包含以下几点：

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。

## 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：

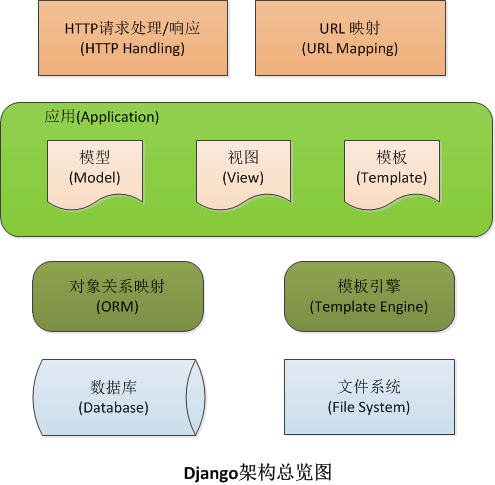


图 3‑5 Django结构概览

### 智能可视化平台

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，

相应的开发任务包括网页开发和移动端适配两部分，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。

echarts图表组件库是一款开源免费的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 3‑6 echarts常见图表展示