

智能文档分析平台

一站式文档处理智能助手，释放文档新活力

AI能力引擎产品部 - 陈刚

目录

01 大模型新质生产力、加速产业智能化变革

02 智能文档分析平台架构概述

03 文档抽取产品特点详解和实操

01

大模型新质生产力、加速 产业智能化变革

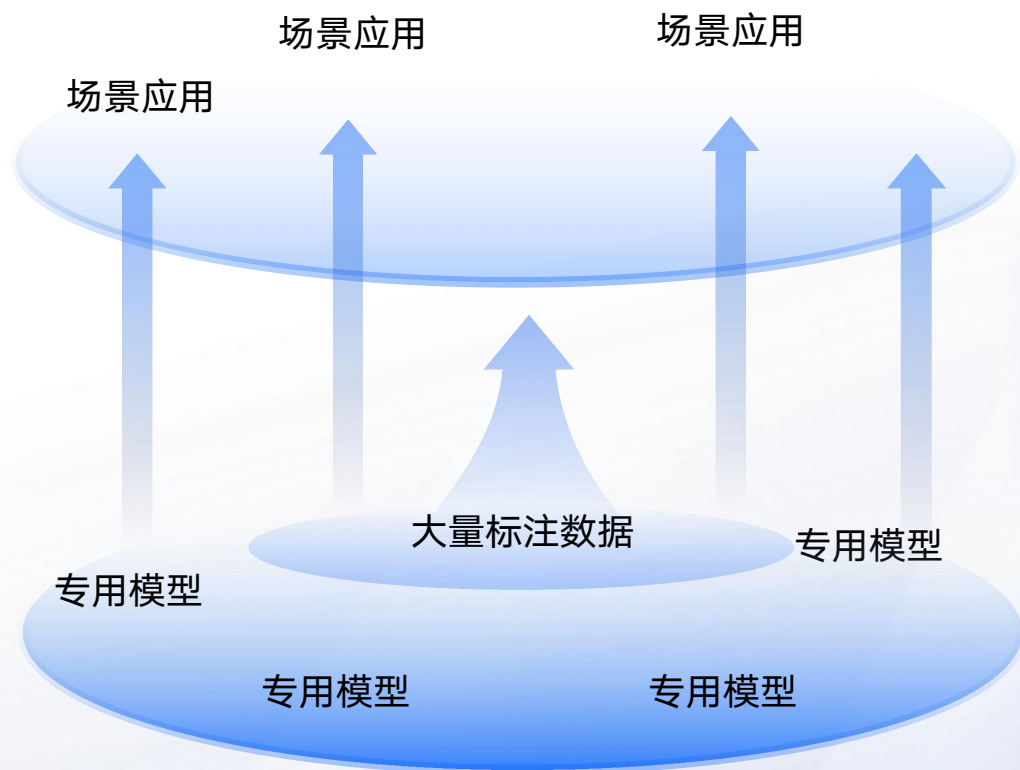
人工智能近年来迎接大模型发展的窗口期，拥变革、创新机

超大规模训练数据、复杂的深层模型和分布式并行训练，造就了这一正在崛起的变革力量



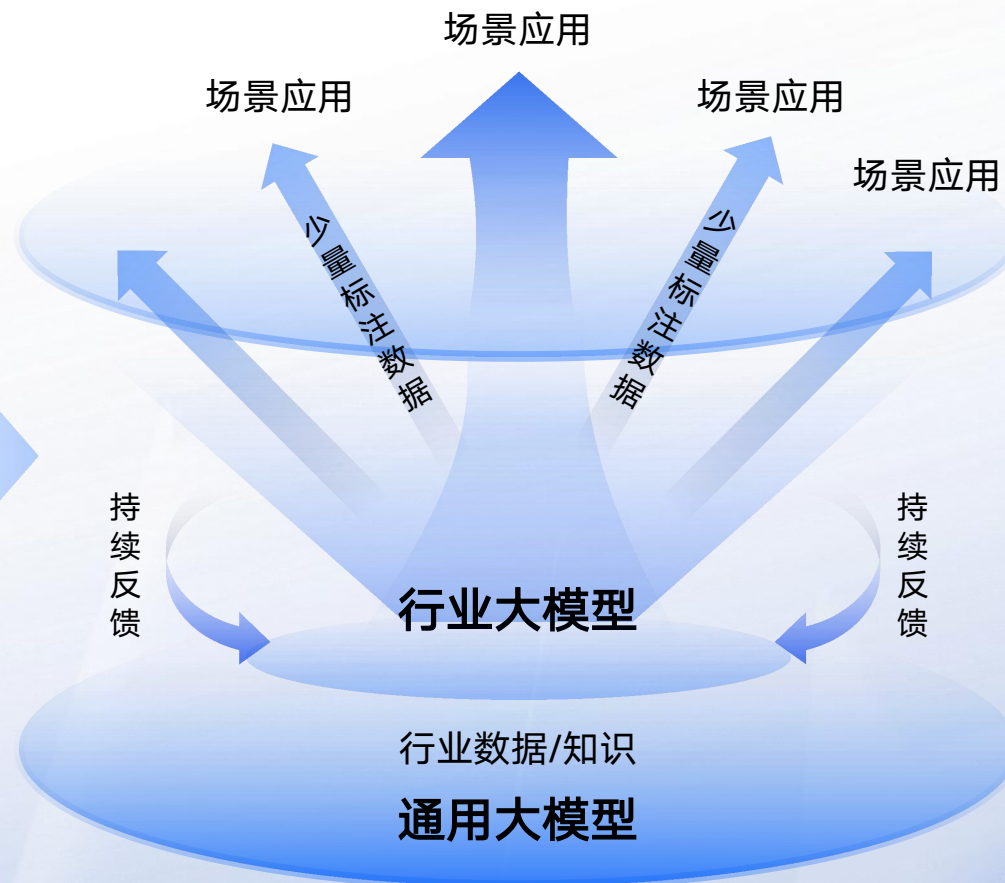
大模型带来产业应用新范式

传统模型应用范式



升级

大模型应用范式



02 智能文档分析平台架构概述

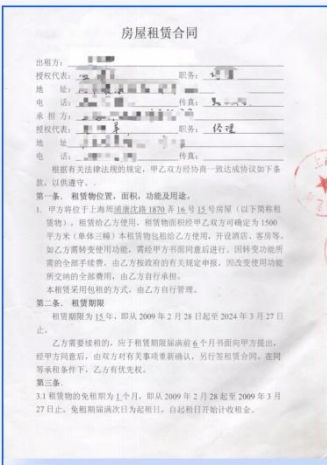
企业文档类型繁多，处理场景多且压力大，急需智能化提效减负



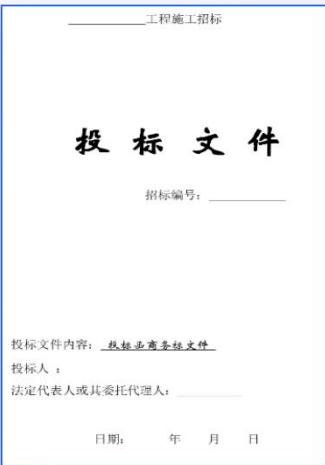
- 01 企业文档“三多一少”：文档种类繁多、文档格式多样、版式布局多样、训练样本稀少
- 02 企业文档多为非固定版式，基于传统AI开发，定制化程度高，训练样本大、周期长，整体成本较高
- 03 基于传统AI开发的文档审查，规则配置复杂，缺乏逻辑推理判断能力

通用文档

合同文书



招投标书

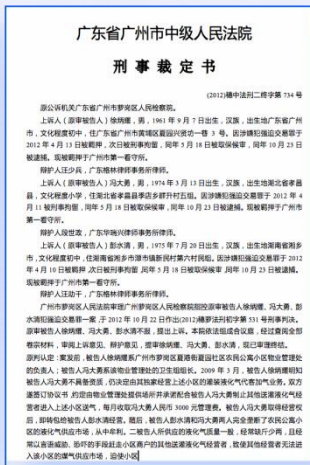


专业文档

保理登记证



法律判决书

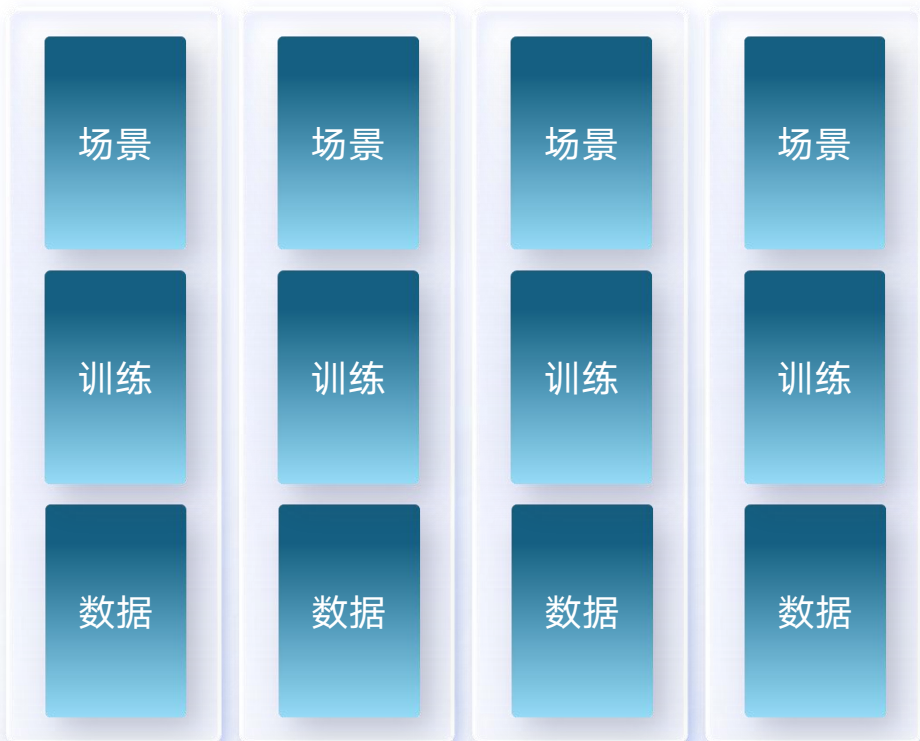


物流单据



大模型对文档智能化处理的革新

基于传统AI的应用开发



基于大模型的文档智能化分析



产品架构图



依托百度文心大模型、OCR等AI技术，为用户提供一站式文档智能化处理服务，帮助企业挖掘文档价值，提升处理效率，加强风险防控

行业实践

泛科技

能源

金融

工业

制造

零售

应用层

合同智能体

招采智能体

文档电子化

金融业务审查

检书业务审查

物流单证校验

能力层

文档抽取

文档比对

合同审查

文档解析

文档版式还原

标书审查

智能文档分析平台 跨模态文档大模型分析引擎

平台层

SaaS 服务（开箱即用）

预置审查 Prompt

Prompt 管理

预置场景模型

PaaS API+SDK（被集成）

数据标注

模型训练&推理

模型管理

多模态布局增强文档模型（ERNIE-Layout）

OCR识别结构化模型（VIMER-StrucText）

文档格式解析（DocParser）

文档版面分析与识别（OCR）

大模型基座层

ErnieBot 4.0 超千亿参数基础大模型

Ernie Speed 百亿参数基础大模型

基础设施层

异构算力

03 文档抽取产品特点详解和实操

文档抽取：无需训练、开箱即用，支持自定义字段

- 定义：基于文心大模型、OCR等技术，无需训练即可自定义抽取词句短语、长段落、表格等多模态内容，解决传统小模型在复杂表格抽取等场景效果受限、泛化性差等问题
- 应用场景：物流单证、金融票据、合同文本、医疗诊单等

无需训练
开箱即用



支持自定义
抽取字段



综合准确
率高



百度智能云 智能文档分析平台

文档抽取

支持自定义配置字段，无需训练即可抽取文档字段信息，精准定位字段值，适用于合同、票据、订单等各类文档场景。如需集成接口调用，可使用文档抽取API服务。

上传文件

报关单示例-3.xls

成功

随附单证2：代理报关委

标记
唛
码
及
备
注
本批货物均为新品，未

项号	商品编号	商品名	数量	单价	总价	币制	原产国	目的地	税率	
1	4016931000	牌子： 容量5L 220V、 型号：								
2	4016931235	牌子：Alpha；规格： 容量2L、额定电压 220V、额定功率50W； 型号：KJQ-6000X	30.00	个	171.20	5,136.00	EUR	西班牙	中国 湖北省 武汉市	照章 征税

报关人员 报关人员证 电话 兹申明对以上内容承担如实申报、依法纳税之法律
责任

申报单位 丁方有限公司 申报单位（盖章） 海关批注及签章

上传文件

*选择文件：

将文件拖拽到此处，或 点击上传

支持上传JPG、JPEG、PNG、BMP、TIFF和TIFF格式图片及DOC、DOCX、PDF、XLSX、XLS和
OFD格式文档。单个文件大小不超过50MB，页数不超过1000页

取消 确定

预置清... 清单3 -- 清单2 -- 新建清单 >>

清单ID: mf-j4da8upufr-6v17

配置抽取字段

单个字段

输入抽取字段后回车确认

组合字段

组合1

输入抽取字段后回车确认

查看抽取结果

保存清单 发起抽取 下载结果

文档抽取：示例图

进出口报关单

中华人民共和国进口货物报关单

预录入编号: 海关编号: 页码/页数:

境内收货人 甲方有限公司	进境类别 广州机场	进口日期 20240725	申报日期 20240727	备案号 C0402500432			
境外发货人 乙方有限公司	运输方式 航空运输	运输工具名称及航次号 CA981	提运单号 COSU123456789	货物存放地 武汉天河机场			
消费使用单位 (914417007718670082)	监管方式 一般贸易	征免性质 一般征税	许可证号 XJ2024070001	启运港 法国			
丙方有限公司	一般贸易	一般征税	XJ2024070001	法国			
合同协议号 HT20240730-001	贸易国(地区) 荷兰	启运国(地区) 德国	经停港 英国	入境口岸 广州白云国际机场			
包装种类 0 纸箱	件数 1	毛重(千克) 1.5	净重(千克) 0.8	成交方式 FOB	运费 5/1421.43/RMB	保费 1/0.0313	杂费

查看抽取结果

启运港
法国

合同协议号
HT20240730-001

保费
1/0.0313

标记唛码及备注
本批货物均为新品, 未使用过。请注意防潮、防震。

申报单位
丁方有限公司

完税证明 - 支持属性组合抽取

中华人民共和国
税收完税证明

No. 3110152407001

纳税人识别号: 911101066738007

纳税人名称: 森林动态有限公司

金额合计: 人民币壹拾肆元玖角柒分

金额合计: ¥14.97

原凭证号	税种	品目名称	税款所属时期	入(进)库日期	实缴(退)金额
007	印花税	经济合同	2024-04-01至 2024-06-30	2024-07-05	14.97

组合1

原凭证号	税种	品目名称
3110162407001	印花税	经济合同

医院处方

麻省儿童医院门诊外配药处方

处方号: 5547007 医保备案号: 外配处方

姓名: 詹姆斯 性别: 年龄: 6岁 门诊号: 80069007 体重: 20kg

家庭地址: 就诊科室: 皮肤科 电话:

初步诊断: 特应性皮炎(中重度)

Rp.	药品名称	单次剂量	频率	途径	份数	发药量	备注
	度普利尤单抗注射液(外配)(2ml: 300mg)	600.00mg	ST	皮下注射	2份	1盒	外配

医师签名:

注: 1. 该处方当日有效, 因特殊情况该处方有效期为____天(签名: _____)
2. 可自行选择药店

委托书 - 支持抽取字段辅助说明

根据我单位经营范围(医疗执业许可范围), 兹委托我单位在职工 张明 同志(身份证号码: 3310819707102465 联系电话: 13325760558) 与贵单位联系购买如下产品:

- 特殊管理的药品: 麻醉药品(包括按麻醉药品管理的品种); 第一类精神药品; 第二类精神药品;
- 国家有专门管理要求的药品: 蛋白同化制剂、肽类激素; 含特殊药品复方制剂(包括含麻黄碱复方制剂、复方地芬诺酯片、复方甘草片); 终止妊娠药品;
- 其他药品;
- 医疗器械;
- 其他产品(食品、化妆品、保健用品等)。
- 毒性药品: 因 治疗 (用途), 需前来购买保妥适(治疗用A型肉毒毒素) 100单位 冻干粉针 美国爱力根爱尔兰, 购买数量 50 支; 保妥适(治疗用A型肉毒毒素) 50单位 冻干粉针 美国爱力根爱尔兰, 购买数量 50 支。

授权委托范围: 代为签署、修改、签收与贵单位之间的《药品购销合同》(或订单)、签收单等全部交易文件, 及以我单位名义处理一切与之有关的事务。

受托人的一切行为, 均将之视为我单位之行为。

授权委托有效期自 2024 年 1 月 1 日至 2024 年 12 月 31 日止。

查看抽取结果

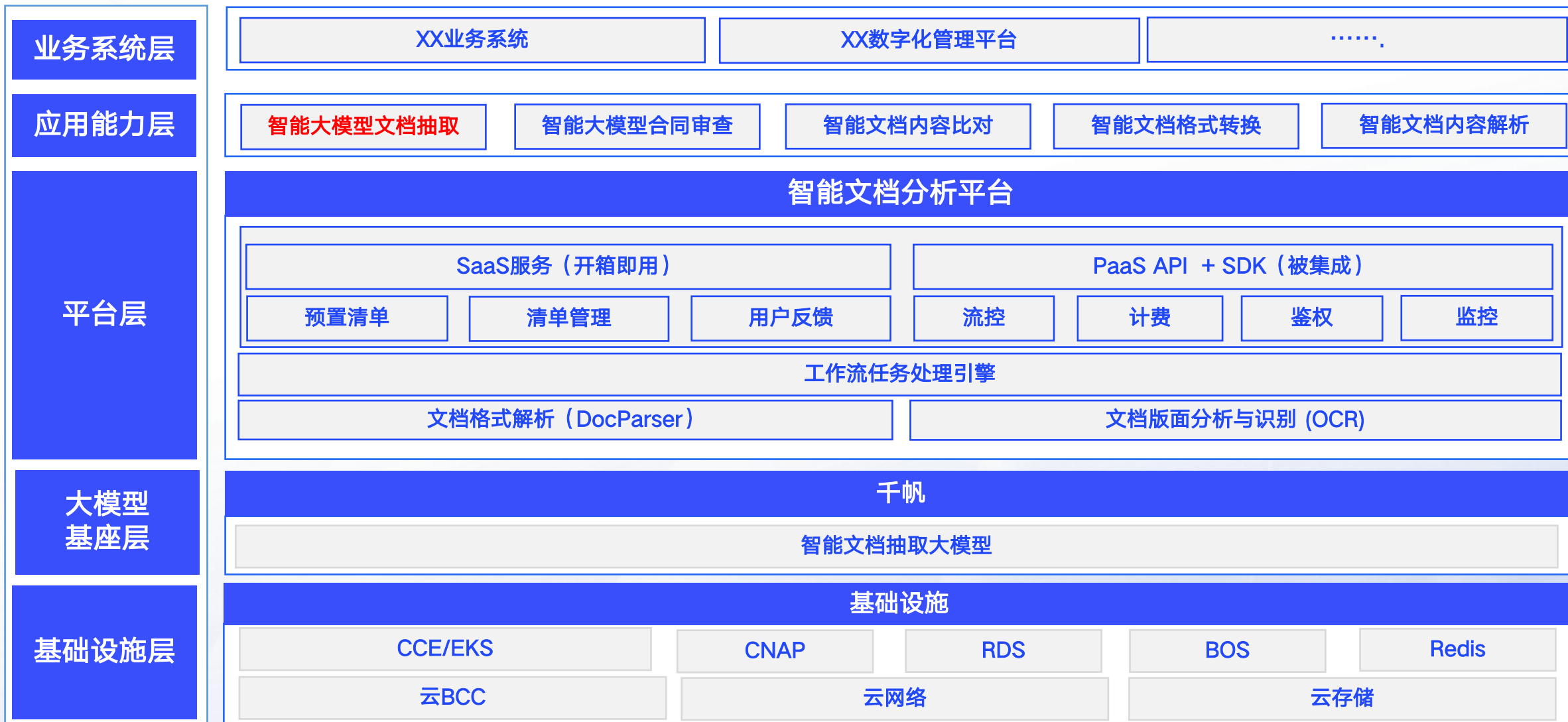
授权委托有效期开始
2024年1月1日

授权委托有效期结束
2024年12月31日

说明

- 授权委托有效期开始: 授权有效期不存在/字符, 只有整数类型
- 授权委托有效期结束: 授权有效期不存在/字符, 只有整数类型

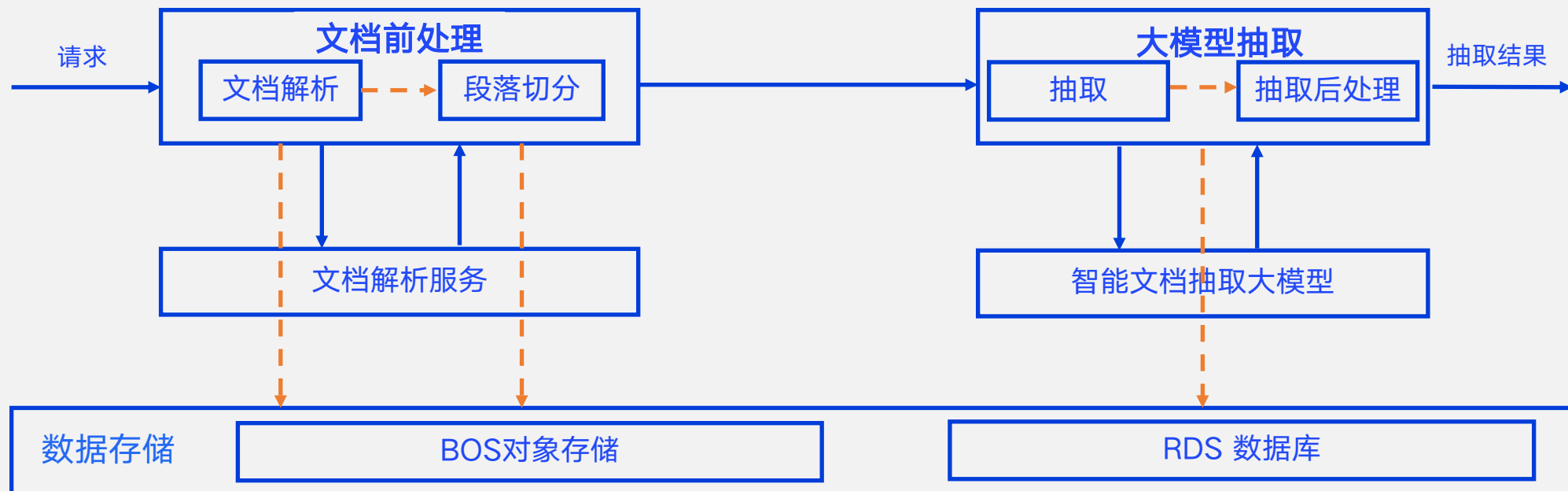
文档抽取 - 公有云产品架构



文档抽取 - 任务处理架构

大模型文档抽取任务处理流程

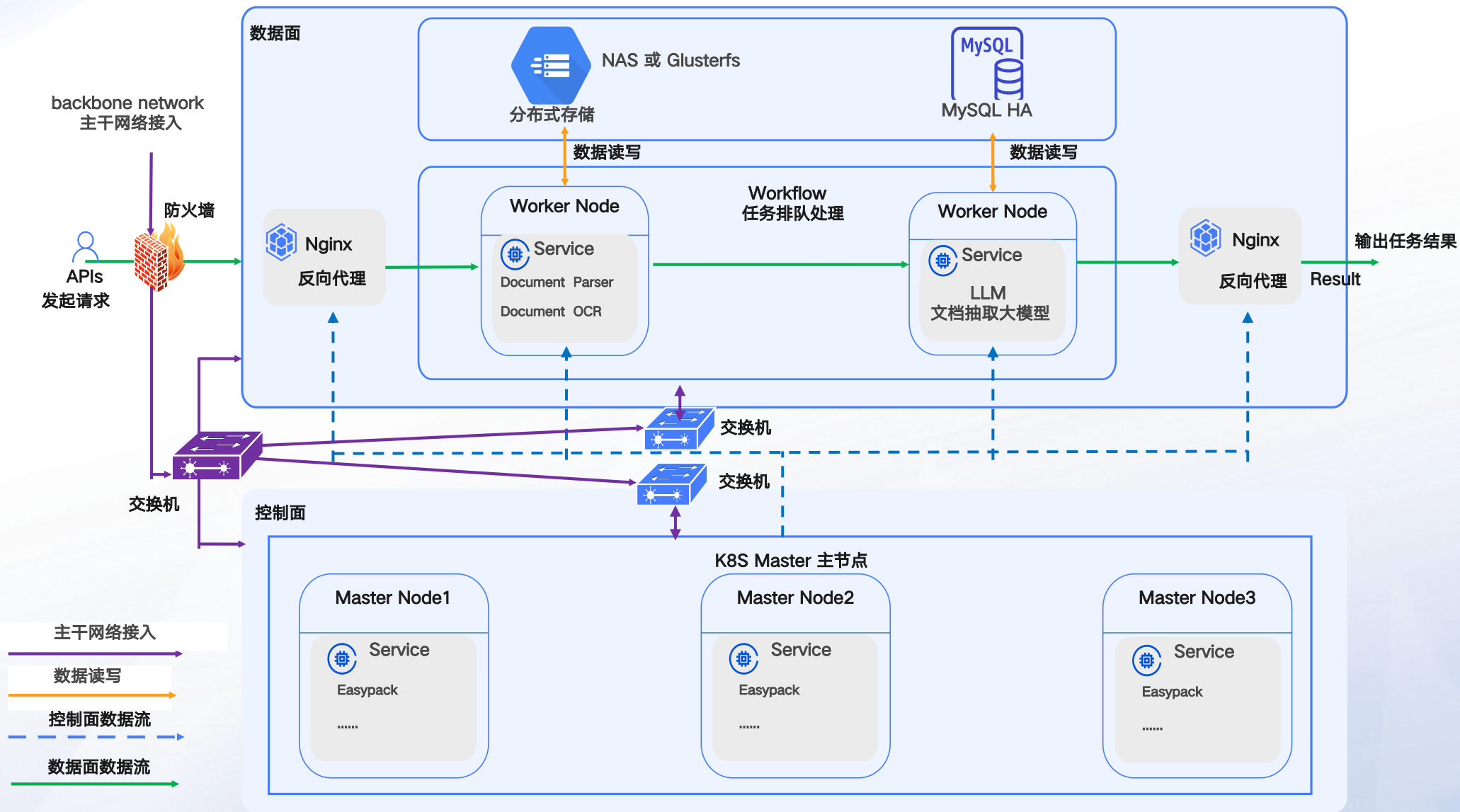
—→ 请求流
- - -> 数据流



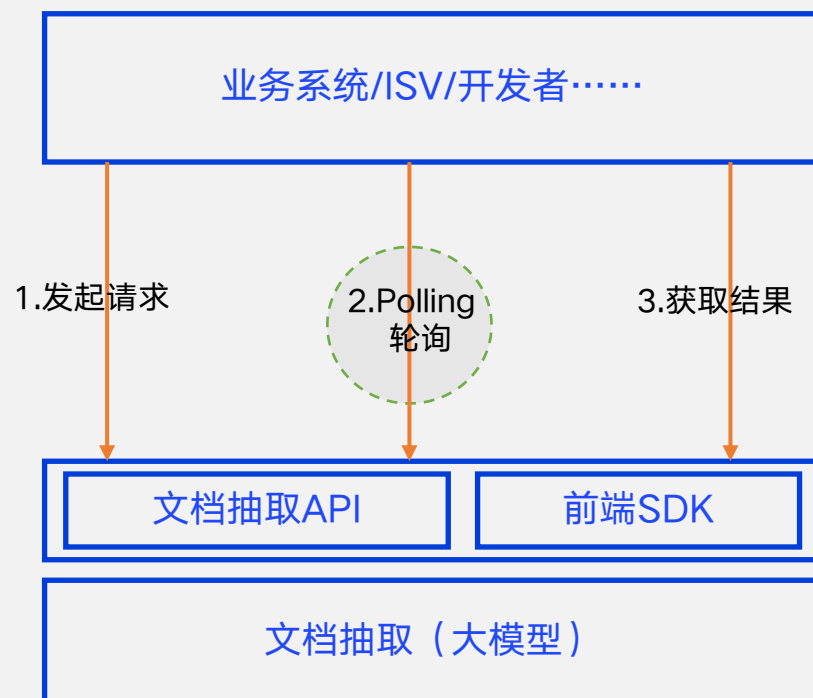
文档抽取 - 私有化产品架构



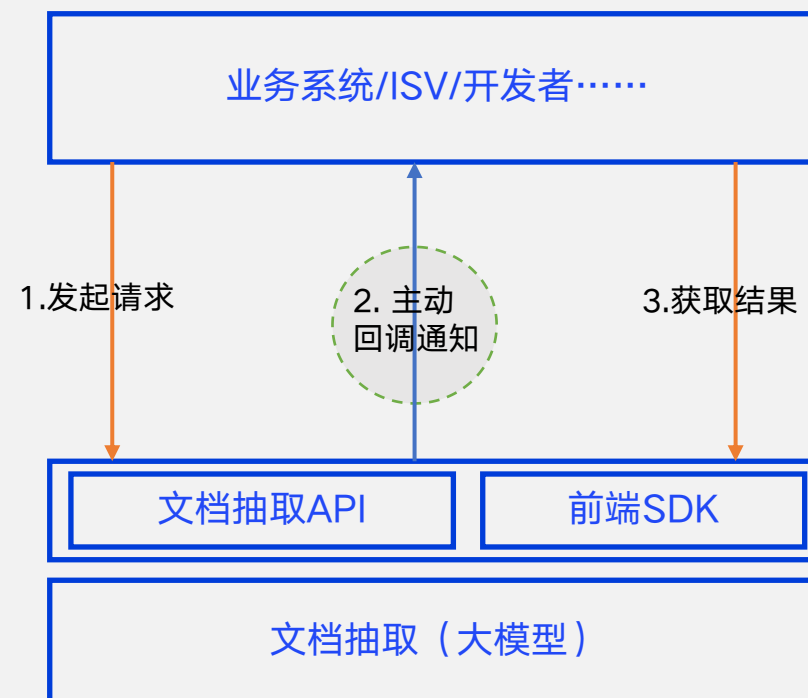
文档抽取 - 私有化技术架构



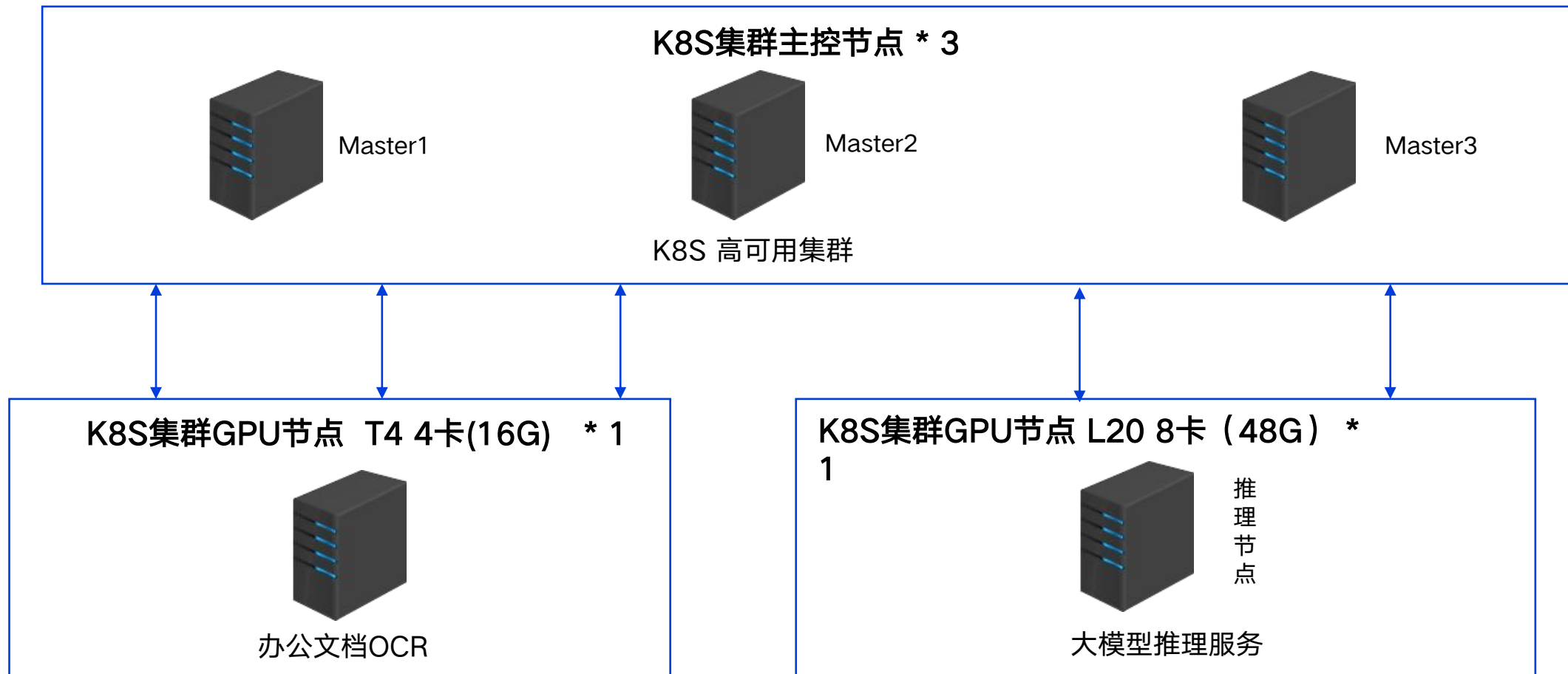
- Pull拉模式：客户/集成商集成为主



- 轻量级推模式：主动回调通知、客户拉取结果（后续上线）



大模型文档抽取 - 纯推理版 - 部署架构 (推荐资源)



支持的硬件、操作系统，**完善且灵活的国产化信创适配**

CPU	GPU	操作系统
<ul style="list-style-type: none">• X86: Intel、海光等• ARM: 鲲鹏系、飞腾系	<ul style="list-style-type: none">• NVIDIA: A100、A800、H800、L20卡等• 昆仑芯: P800• 华为: 910B系	<ul style="list-style-type: none">• Ubuntu• CentOS• 麒麟系

制造业案例

应用于某制造业的合同管理系统和财务系统，对合同和财务票据数据进行关键信息抽取

出行案例

应用于某地出行科技公司，处理非机动车三方责任险保单数据，对投保人、被保险人等主要信息进行结构化提取

基于行业应用实践，持续沉淀场景方案

- 助力能源、金融、物流、政务等行业客户降本增效，巩固风险防控，已在100+头部企业落地实践



典型应用方案

合同智能审查

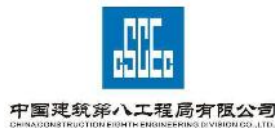
贸易物流单证校验

智慧招采

金融业务审查

检书业务审查

文档抽取





问答环节

