联想全栈AI
加速智能化转型每一步

联想智能云（Lenovo xCloud）产品介绍

模型、智能体、数据与知识开发一体机 & 工作站

联想智能云全景图

联想沉淀40年IT智慧，以混合云为坚实底座，持续打造企业数据和知识库、模型工厂、智能体平台以及人工智能安全体系，助力客户在AI时代稳健前行。

联想智能云（Lenovo xCloud）

模型平台

模型训练及客制化	模型推理	模型仓库
大模型微调与蒸馏方案 模型开发平台（训推版） 模型开发一体机/工作站 炼AI大师AI Master	模型开发平台（推理版） 模型开发工作站（标准版） 模型开发工作站（基础版） 模型开发一体机（标准版）	企业模型库 模型库开发一体机

智能体平台（AI Force）

智能体开发
智能体开发平台 智能体开发一体机/工作站

数据和知识平台

数据合成	知识管理
语料库建设与数据标注方案	企业知识库 知识蒸馏 知识治理 知识库开发一体机/工作站

混合云

SAP云	私有云	KT云	云原生	云一体机	智算云	AIOps	多云管理&FinOps
SAP HANA系统 SAP PE CDC SAP备份系统 SAP实施与迁移 SAP应用系统 SAP业务应用方案	IoT网关 私有云平台 裸金属管理 云建设&云迁移方案	KT AIOps KT DCM KT私有云 KT云原生 KT云原生一体机 KT私有云一体机	DevOps 数据库及中间件管理 容器平台 微服务治理 云原生现代化解决方案	云原生一体机 私有云一体机	智算云平台	备份容灾 数据中心管理DCM 智能运维(AIOps) 方案	FinOps 多云管理平台 云安全与合规

智算中心

基础配套	一体化机房	智能动环监控
基础配套 智算中心设计与建设方案	一体化机房	AI DCM

TruScale算力服务

AIDC aaS
数据中心机柜订阅 联想数据中心设备订阅 第三方数据中心设备订阅

智能云托管服务
AI IaaS 联想企业云 (LEC) 公有云MSP服务 模型即服务（MaaS）

金融服务

AI相关开发工作站 /一体机

模型开发一体机-训推

客户场景

- 基于DS模型微调，蒸馏等需求
- AI一体机支持DS满血版部署

模型开发（训推）一体机 | 模型开发（训推）工作站

联想问天 WA5480G3
8*NV L20

联想问天 WA7880a G3
8*沐曦 C550

ThinkStation PX
4*NV RTX5000/5880
4*MetaX N260

模型开发一体机-推理

客户场景

- 需要快速部署DS模型
- 利用模型进行推理，无需微调

模型开发（推理）一体机 | 模型开发（推理）工作站

联想问天 WA5480G3
4*NV L20
4* MetaX N260

ThinkStation P5
2*NV RTX5000
2* MetaX N260

ThinkStation P620
4*NV RTX4000
2* MetaX N260

知识库开发一体机

客户场景

- 实现知识库无缝整合，统一管理
- 提供智能问答，提升知识管理与利用效率

知识库开发一体机 | 知识库开发工作站

联想问天 WA5480G3
2*NV L20
2*Meta X N260

ThinkStation P5
2*NV RTX4000
2*Meta X N260

ThinkStation P620
2*NV RTX4000
2* Meta X N260

智能体开发一体机

客户场景

- 基于大模型快速开发智能体应用
- 支持拖拽操作构建智能体网络

智能体开发一体机 | 智能体开发工作站

联想问天 WA5480G3
4*NV L20 /8*NV L20
4*Meta X N260 /
8* Meta X N260

ThinkStation PX
4*NV RTX5000
4* Meta X N260

联想全栈AI

加速智能化转型每一步

模型开发工作站&一体机（推理版）

联想40周年

模型开发一体机/工作站 方案一览表

分类	模型开发一体机/工作站	
版本	模型开发平台（推理版）	模型开发平台（训推版）
MA	Per node w/1年MA	Per node w/1年MA
用户规模	几十到几百用户的B3客户或B2客户部门级	几百用户的B2客户
算力芯片	工作站：ThinkStation P3/P5/P620，RTX 5000 / RTX4000 / 沐曦 N260 服务器：WA5480/WR5220，L20/沐曦C500/沐曦260	工作站：ThinkStation PX，RTX 5880 / RTX5000 / 沐曦 N260 服务器：WA7880a/WA5480，L20/沐曦C550/沐曦N260
规模	1节点	1节点、4节点
模型应用	私有化大模型API服务、本地知识问答	私有化模型训练、满血版模型本地API服务
模型	DeepSeek R1、Qwen 3、Gemma 3	DeepSeek R1、Qwen 3、Gemma 3
模型服务	本地高速、无限Tokens的大模型推理服务	本地高速、无限Tokens的大模型推理服务 专属领域、企业专属模型微调
增值功能	多模型切换；新模型部署	身份认知微调；模型评估
服务	模型应用咨询、培训；模型对接	数据采集、标注服务；模型微调

模型开发一体机/工作站（训推版）

产线：联想智能云（Lenovo xCloud
产品经理：刘志恒
13521976746

产品定位：适用于有基于 DS 模型进行微调及蒸馏需求的用户，高配机型支持 DeepSeek 满血版的高效部署

方案组成

软件平台

联想智能云（Lenovo xCloud）模型开发平台-专业版(模型训推)

数据工程	模型管理	微调训练	评估优化	部署推理
数据管理	模型集市	指令工程	模型评测	在线推理
数据标注	我的模型	参数微调	模型比对	离线推理
数据回流	模型体验	分布式训练	模型量化	分布式推理
数据增强		指标可视化	模型蒸馏	多Lora推理
数据处理		异构兼容		调用监控

大模型：DeepSeek/Qwen/GLM...

算力平台

模型训推AI工作站
ThinkStation PX
4*NV RTX5000 / 4*NV RTX5880 / 4*沐曦N260

模型训推AI一体机
WA7880a G3　8*N沐曦 C550
WA5480G3　8*NV L20

方案亮点

零代码模型精调

- 无需编写代码，轻松定制专属模型
- 断点续训技术，训练过程无忧

自动模型评测

- 多指标量化，精准评估模型性能
- 全维度对比，快速锁定最优模型

一键推理部署

- 一键部署，高效上线推理服务
- 实时响应，快速输出推理结果

模型开发一体机/工作站（推理版）

产线：联想智能云（ Lenovo xCloud
产品经理：刘志恒
13521976746

产品定位：面向需要快速本地部署DeepSeek、Qwen3等主流大模型，直接利用模型进行应用，助力构建本地的安全模型推理能力

方案组成

软件平台

联想智能云（ Lenovo xCloud ）模型开发平台-标准版(模型推理)

模型管理	评估优化	部署推理
模型集市	模型评测	在线推理
我的模型	模型比对	离线推理
模型体验	模型量化	分布式推理
	模型蒸馏	多Lora推理
		调用监控

大模型：DeepSeek/Qwen/GLM...

算力平台

模型开发工作站

ThinkStation P5：2*NV RTX5000 / 2*沐曦 N260

ThinkStation P620：4*NV RTX4000 / 2*沐曦 N260

模型开发一体机

WA5480G3

4*NV L20、4*沐曦 N260

方案亮点

高效推理加速

- 模型量化技术，保证精度，降低消耗
- 集成高效推理引擎，加速推理过程

灵活部署管理

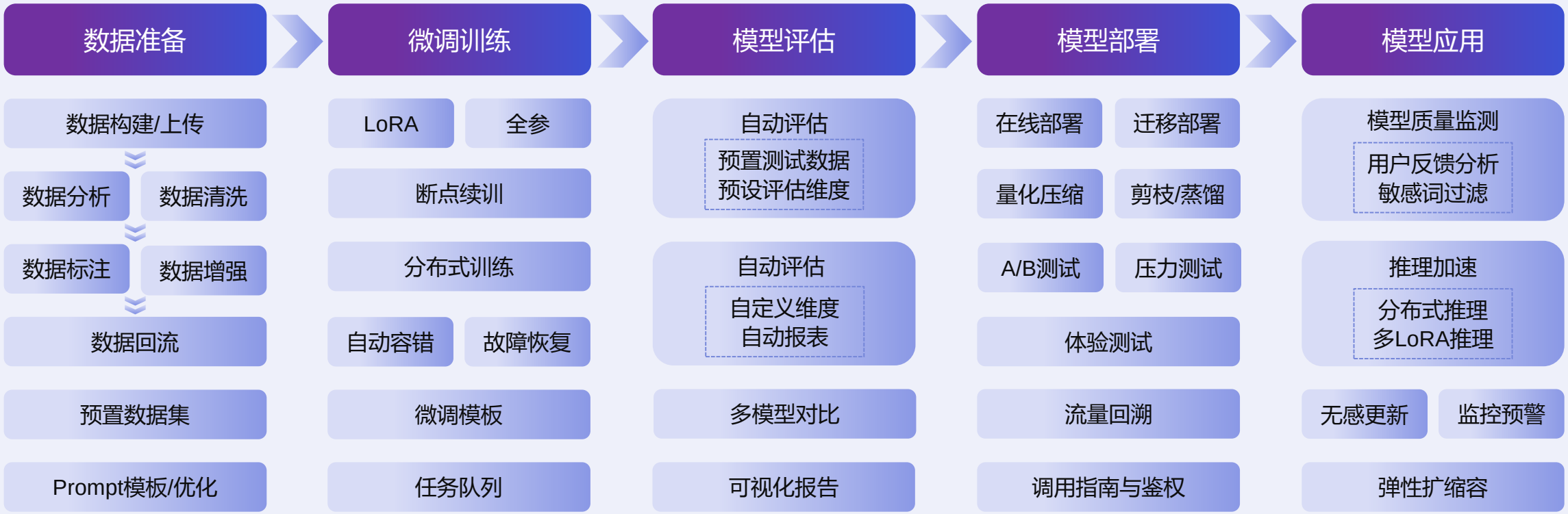
- 多种部署集成方案，无缝衔接业务系统
- 灵活扩缩容能力，弹性调整资源占用

智能运维监控

- 多维度全景监控，实时感知推理性能
- 自动故障感知，模型服务稳定可靠

模型开发平台：提供大模型全链路工具服务

针对大模型开发过程中的**数据处理难、开发成本高、模型落地难**等痛点，提供大模型全链路工具，全方位降低模型开发门槛，实现大模型全生命周期的闭环管理和模型迭代优化。



联想全栈AI

加速智能化转型每一步

智能体开发
工作站&一体机

Lenovo

智能体开发一体机/工作站 方案一览表

分类	智能体开发一体机/工作站		
版本	智能体开发平台（标准版）	智能体开发平台（专业版）	智能体开发平台（旗舰版）
MA	Per node w/1年MA	Per node w/1年MA	Per 套 w/1年MA
用户规模	几十用户的B3客户或B2客户部门级	几十用户的B3客户或B2客户部门级	几十用户的B3客户或B2客户部门级
算力芯片	工作站：ThinkStation PX，RTX 5880 / RTX5000 / 沐曦 N260 服务器： WA7880a/WA5480/WR5220，沐曦C550/沐曦C500/L20	工作站：ThinkStation PX，RTX 5880 / RTX5000 / 沐曦 N260 服务器： WA7880a/WA5480/WR5220，沐曦C550/沐曦C500/L20	工作站：ThinkStation PX，RTX 5880 / RTX5000 / 沐曦 N260 服务器： WA7880a/WA5480/WR5220，沐曦C550/沐曦C500/L20
规模	1节点	1节点、2节点	无限节点
模型应用	私有化大模型API服务、本地知识问答	私有化模型训练、满血版模型本地API服务	
模型	DeepSeek R1、Qwen 3	DeepSeek R1、Qwen 3	DeepSeek R1、Qwen 3
基础功能	智能应用编排能力、知识库挂载	多模态能力、意图分类引擎、提示词优化	长期记忆、智能体网络、数据库接入
增值功能	N/A	智能体发布API接口/SDK	模型微调
服务	智能体开发咨询、培训；智能体定制开发	智能体开发咨询、培训；智能体定制开发	智能体开发咨询、培训；智能体定制开发

智能体开发一体机/工作站

产品定位：基于大模型快速开发智能体应用，支持拖拽操作来构建智能体网络，大幅降低智能体应用开发的门槛与难度

方案组成

软件平台

智能体开发平台AI Force

智能体样例	合同审批助理	文档速览大师	信审助手	规文笔杆	销售顾问	数据分析专家

提示词优化	检索增强	插件	配置编排	评估管理	部署发布	运营监控	模型微调
模板库	文档标注	插件注册	调试器	模型评估	版本管理	行为分析	数据处理
优化器	智能分片	插件封装	场景配置	知识评估	场灰度发布	安全监控	模型训练
提示扩写	向量化	插件测试	可视化编排	应用评估	权限控制	报表/大屏	模型压缩

大模型：DeepSeek/通义千问

算力平台

智能体开发工作站
ThinkStation PX
4*NV RTX5000
4*沐曦N260

智能体开发一体机
WA5480G3
4*NV L20 / 8*NV L20
4*沐曦N260 / 8*沐曦N260

方案亮点

一体化集成

- 支持丰富大模型，最高支持72B
- 零代码开发工具，分钟级创建

私有化部署

- 数据不出门，安全有保障
- 企业系统快速对接，成果易推广

简易运维

- 桌面级安装
- 7 * 24小时不间断运行

联想全栈AI

加速智能化转型每一步

知识库
开发工作站&一体机

Lenovo

知识库开发一体机/工作站 方案一览表

分类	知识库开发一体机/工作站
软件版本	企业知识库
MA	Per 套 w/1年MA
用户规模	几十到几百用户的B3客户或B2客户
算力设备	工作站：ThinkStation P5/P620/PX，RTX 5000Ada / RTX 4000Ada / 沐曦 N260 服务器：WA5480, L20/沐曦C500/沐曦260
规模	1节点、2节点、4节点
应用场景	本地智能知识管理、知识问答、API对接
模型	DeepSeek R1、Qwen 3、Qwen-VL、BAAI/bge-m3、BAAI/bge-reranker-large
模型服务	本地高速、无限Tokens的大模型推理服务，数据不出域
增值功能	知识运营；多模态数据解析、时效性&重复性双重检测
服务	知识库构建、知识优化

知识库开发一体机/工作站

产线：联想智能云（Lenovo xCloud
产品经理：刘志恒
13521976746

产品定位：助力企业实现内部知识库的无缝整合，提供统一管理功能，并支持智能问答，提升企业知识管理与利用的效率

方案组成

软件平台

联想智能云（Lenovo xCloud）企业知识库

知识导入	知识管理		知识运营
多源知识导入	知识权限	知识同步	知识反馈调优
知识切片	知识标签	问答集抽取	知识统计
向量索引	重复度检测	时效性检测	知识运营

大模型服务

RAG | 逻辑推理 | 知识检索 | 模型服务

大模型：DeepSeek/Qwen/GLM...

算力平台

知识库开发工作站

ThinkStation P5：2*NV RTX4000 / 2*沐曦 N260

ThinkStation P620：2*NV RTX4000 / 2*沐曦 N260

知识库AI一体机

WA5480G3

2*NV L20、2*沐曦N260

方案亮点

智能知识治理

- 异构知识采集管理，多模态解析
- 时效性&重复性双重检测，确保知识鲜活

语义精准理解

- 自动识别知识内容特性，智能语义分割策略
- 综合重排序模型等排序算法，知识精准输出

数据安全可靠

- 本地化部署，数据不出门
- 知识权限安全管控，保障知识安全

联想智能云 AI 技术方案服务清单

服务分类	服务名称	服务定义	典型适用场景	服务内容及步骤	标准人天	核心交付物
部署实施	Lenovo xCloud软件+DS 部署服务	在客户本地环境安装并启动DeepSeek，使其具备基础功能	首次部署的中小企业客户	硬件环境合规性检查、私有化部署包安装、基础功能联调测试等	N+2	部署手册 部署验收报告
	系统集成	将AI能力嵌入或调用客户现有业务系统	通过集成AI能力赋予业务系统智能决策、自动化执行等能力	通过API对接等与业务流程适配，将AI能力嵌入业务系统（ERP等），实现 数据流-决策流-执行流 的端到端闭环	按需	需求分析 技术设计方案 测试报告与验收 用户手册与培训材料
	模型微调	基于客户私有数据对模型进行微调，提升特定任务表现	适配行业术语与业务流程，领域知识突破原生限制	轻量化微调、知识注入、效果验证等	按需	配置文件与参数说明 微调效果对比评估
运行维护	模型升级服务	版本迭代的平滑过渡保障	新模型版本发布后，应用或验证新特性	版本兼容性测试、增量升级实施、回滚方案准备等	按需	升级记录 新功能文档
	运维服务	帮助客户解决在使用模型过程中遇到的问题	模型训练，推理，应用过程中遇到问题	解决在使用DS等模型过程中遇到的配置、性能、兼容性相关问题	按需	问题分析手册 RCA报告
	推理性能优化	模型推理效能专项优化	提升高并发、低延迟需求的实时AI应用及资源受限环境中的模型推理效率。	推理加速优化、内存压缩、硬件加速配置等	按需	性能报告 基准测试数据

Lenovo 联想

thanks.

联想40周年