JISUMEN | JISUMEN Technology

Get AI running.
Manage it day to day.

Deliver compute, models and business systems as enterprise AI capabilities that are usable, verifiable and able to evolve.

Enterprise Private AI Infrastructure and Implementation Services

COMPANY PROFILE 2026



AI in real business demands a new definition of infrastructure

Enterprises need more than a device or a model. They need a platform they can operate, validate, and improve within their own data boundaries.

Controllable

Enterprises retain control of compute, models and data boundaries.

Usable

Capabilities reach real work interfaces, beyond a demonstration environment.

Evolvable

The architecture evolves with models, workloads and business needs.

This is where JISUMEN's new understanding of delivery begins.

JISUMEN: your private AI engineering partner for real business

Engineering connects liquid-cooled AI workstations, private compute, model operation and enterprise AI applications.

JISUMEN
极算门

Registered trademark | No. 89402856 | Class 9

Dongguan JISUMEN Computer Technology Co., Ltd.

We serve enterprises, education and research institutions with liquid-cooled AI workstations and private AI infrastructure, model deployment, application integration and ongoing operations.

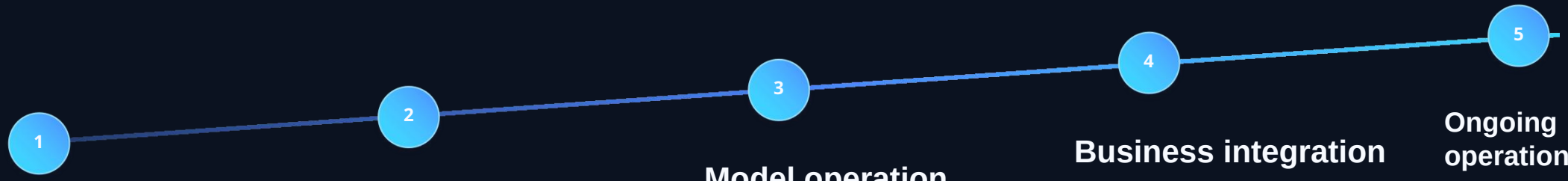
Starting with real workloads, we organize hardware, software and business systems into an integrated capability for long-term operation.

Brand expression

The continuous waveform expresses the flow of compute, data and models. JISUMEN and its Chinese name form the brand identity, conveying engineering that bridges the gap to practical AI adoption.

A complete chain connects compute to business value

JISUMEN organizes fragmented technology into a delivery system with defined boundaries, validation and continuity.



Requirements

Business goals
Data boundaries

Infrastructure

Compute and systems
Cooling and networking

Model operation

Deployment and tuning
Service interfaces

Business integration

Agent orchestration
Systems and workflows

Ongoing operations

Validation and monitoring
Upgrades and maintenance

Without any link in this chain, AI is difficult to run reliably in production.

Engineering determines whether AI creates lasting value

Products are the vehicle; cross-layer coordination and continuous delivery make the difference.

Infrastructure engineering	Design compute, power, cooling, networking and storage as one system.	Stability
Models and runtimes	Choose the runtime architecture around models, precision, context and concurrency.	Efficiency
Application and system integration	Bring agents into documents, knowledge and workflows across business systems.	Coordinatior
Validation and ongoing operations	Support results with reproducible tests, acceptance documents and long-term maintenance.	Evolution

A team that understands hardware, models and application boundaries reduces gaps in responsibility between vendors.

The same capabilities, adapted to different scales and stages

Architectures follow workloads, data boundaries and evolution paths, rather than simple product tiers.

Department

Professional teams and R&D

High-performance workstations and local development
Focused on fast starts and professional workloads

Enterprise

Shared access across teams; private deployment

Multi-node inference and unified model services
Focused on collaboration, permissions and internal reuse

Data center

High concurrency and critical business

Data-center AI engineering and performance acceptance
Focused on scale, stability and long-term evolution

Shared principle

Base resource configuration on model and business requirements, with room for future expansion.

Two types of systems, built on the same private AI capabilities

Core delivery standards stay consistent, from desktop multi-GPU compute to team-scale private infrastructure.



liquid-cooled AI workstations

Single / dual / four GPUs

Local large models
AI development and generative AI
Stable sustained workloads

Start quickly at the desktop



Private AI infrastructure

Single machine / multiple nodes

Model deployment
Agent and system integration
Validation and ongoing operations

From compute delivery to operations

Multi-GPU architecture

Full water cooling

Private model deployment

Continuous delivery support

High-density compute in a desktop system built to keep running

Water cooling is system engineering for sustained AI workloads, beyond a cooling method.



Actual internal structure | Dual-GPU full water cooling

JISUMEN liquid-cooled AI workstation

For local large models, generative AI and professional rendering, multi-GPU architecture, CPU/GPU water cooling and system-wide thermal management support stable, long-duration high-density compute.

From hardware selection and custom water cooling to software environments and model deployment, we deliver a system ready for work.

Engineering approach

Water cooling is a complete system design around power, thermal density, expansion and maintenance. Model size and workloads determine the configuration.

Bring AI from answering questions into enterprise workflows

AI agents are the business orchestration layer within private deployment and delivery, rather than a separate product line.

It connects documents, business systems, private models and compute, enabling AI to understand tasks, use tools and complete workflows within enterprise permissions.

01 Business scenarios

02 **Agent orchestration**

03 Private models

04 Compute and operations

Document automation

Collect, recognize, review, generate and archive materials.

Cross-system coordination

Reduce copying, exporting and repeated data entry.

Closed-loop data workflows

Query, process, write back and track within permission boundaries.

Start with a high-value use case and build reusable platform capabilities

Validate value through real problems, then standardize connectors, workflows and operations step by step.

01

Use-case validation

Start with common business documents, such as quotes, contracts, reports, acceptance documents, and meeting notes.

Prove value first

02

System integration

Connect OA / ERP / CRM, file servers, databases and APIs.

Close the workflow loop

03

Platform development

Productize shared connectors, workflow templates, model routing, monitoring and permissions.

Make it repeatable

Especially suitable for small and medium enterprises seeking a low-cost start, gradual expansion and clear data boundaries.

Turn AI project uncertainty into a manageable delivery process

Every stage has clear inputs, outputs, validation criteria and a basis for decisions.

01

DISCOVER

Business and boundaries

Goals, data, models and context Concurrency and budget boundaries

02

DESIGN

Architecture and standards

Compute, systems, runtimes, integration methods and acceptance criteria

03

DELIVER

Deployment and acceptance

Installation, integration, testing, documentation and delivery to real work interfaces

04

EVOLVE

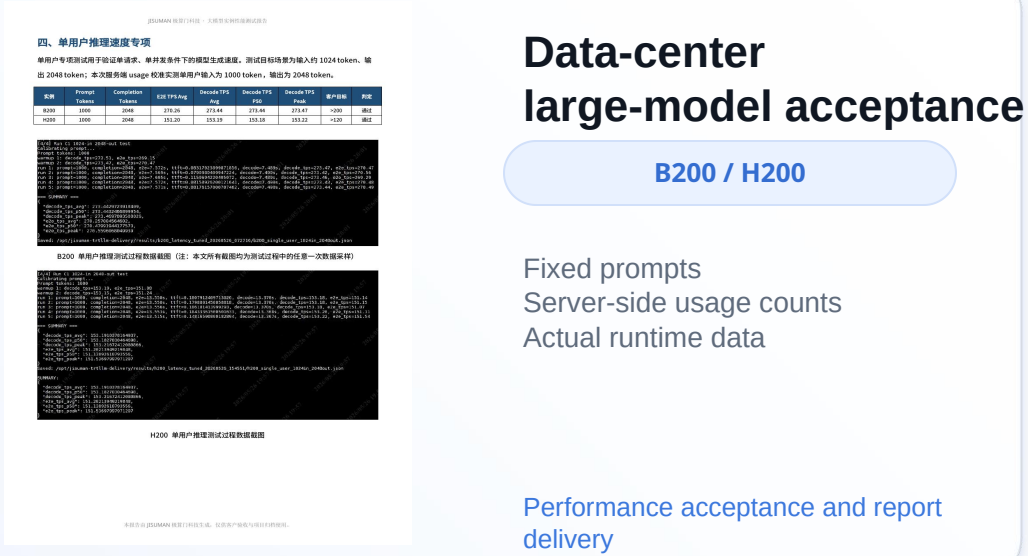
Operations and evolution

Monitoring, optimization, upgrades, expansion and continued business integration

Delivery means reliable use in real work, with clear validation and ongoing maintenance, beyond simply installing a system.

Representative validation: different scales, consistent standards

Devices and models vary by scenario; real environments, explicit measurement methods and reproducible results remain constant.



四、单用户推理速度专项

单用户专项测试用于验证单请求、单并发条件下的模型生成速度。测试目标场景为输入约 1024 token、输出 2048 token；本次服务端 usage 校准实测单用户输入为 1000 token，输出为 2048 token。

实例	Prompt Tokens	Completion Tokens	E2E TPS Avg	Decode TPS Avg	Decode TPS P50	Decode TPS Peak	客户目标	判定
B200	1000	2048	278.26	273.44	273.44	273.47	>200	通过
H200	1000	2048	151.20	153.19	153.18	153.22	>120	通过

B200 单用户推理测试过程数据截图（注：本文所有截图均为测试过程中的任意一次数据采样）

H200 单用户推理测试过程数据截图

本报告由 JISUMEN 极智门科技生成，仅供客户验收与项目归档使用。

Data-center large-model acceptance

B200 / H200

Fixed prompts
Server-side usage counts
Actual runtime data

Performance acceptance and report delivery



Multi-node private compute validation

Single machine / multiple nodes

Model deployment
Cluster integration
Concurrency and boundary validation

Reproducible engineering conclusions

Real environments

Explicit measurement methods

Reproducible results

Deliverable documentation

Choosing JISUMEN means more than choosing equipment

What matters is an engineering system that supports enterprises from their first deployment to long-term operation.

01

Design architecture from real workloads

Understand models, data and business before deciding on compute, nodes, cooling and systems.

02

Define delivery with validation results

Document metrics, parameters, boundaries and raw results to reduce vague promises.

03

Plan ongoing operations from day one

Plan monitoring, maintenance, upgrades, expansion and future business integration from the design stage.

We handle the engineering complexity so customers gain clarity, stability, and a system they can maintain over time.

Long-term partnerships rest on real capabilities and continued investment

Brand, engineering assets, test environments and service systems form a foundation of trust.

Registered brand

JISUMEN

No. 89402856 | Class 9
Valid through June 27, 2036

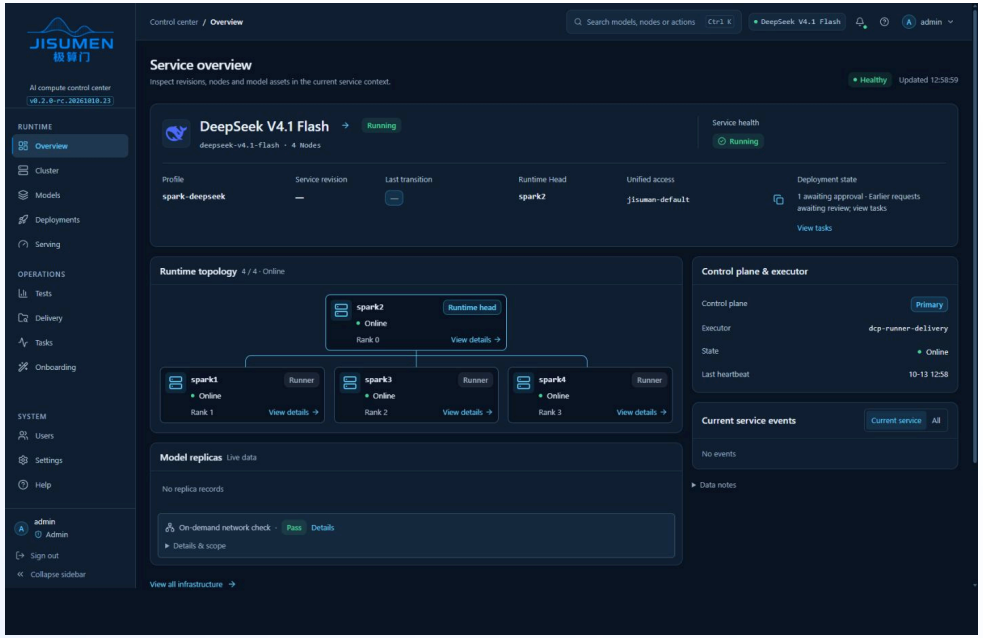
Engineering assets

Real test environments and open-source tools

Continuously building capabilities in model deployment, performance validation, delivery documentation and historical result management.

Service foundation

Dongguan service base | On-site support | Nationwide remote collaboration



Real runtime environments | Model services, tool access and application integration

Continued investment turns project experience into organizational capability

Turn validated engineering experience into stable products, processes and service systems.

01 **Productization**

Define clear solution boundaries for liquid cooling, workstations, private compute and delivery components.

02 **Delivery system**

Standardize requirements assessment, architecture, implementation, acceptance and ongoing operations.

03 **Teams and service**

Strengthen customer support, operations, materials management and assembly for more reliable delivery.

The goal is more stable, clearer and repeatable delivery, rather than a longer product list.

Start with a clearly defined business problem

We first understand five things

01 Business goals

02 Data boundaries

03 Existing systems

04 Models and scale

05 Delivery requirements

Exceptional compute. Gateway to intelligence.

From private compute and model deployment to AI agents, make AI a real enterprise capability.