

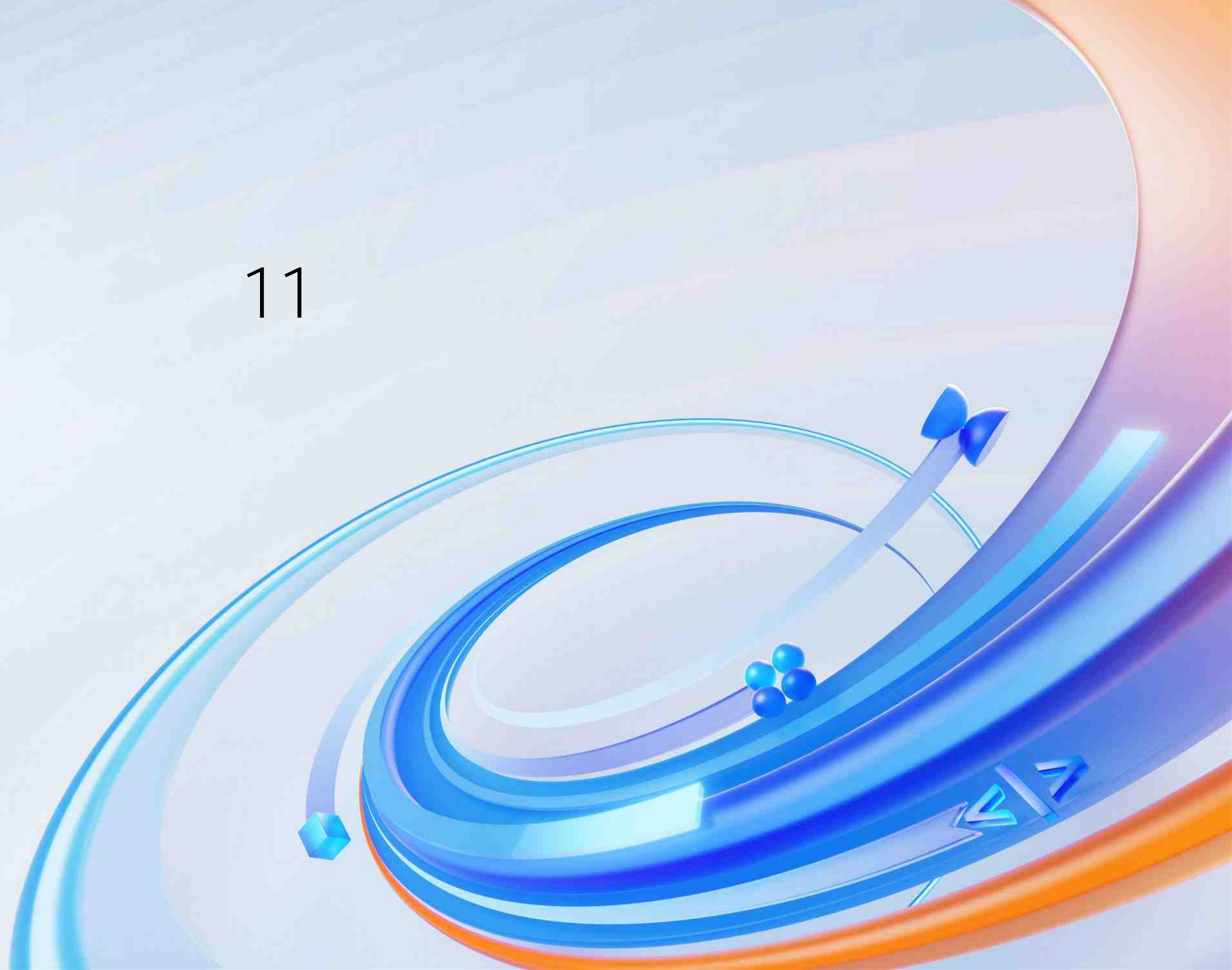
AI

11

AI

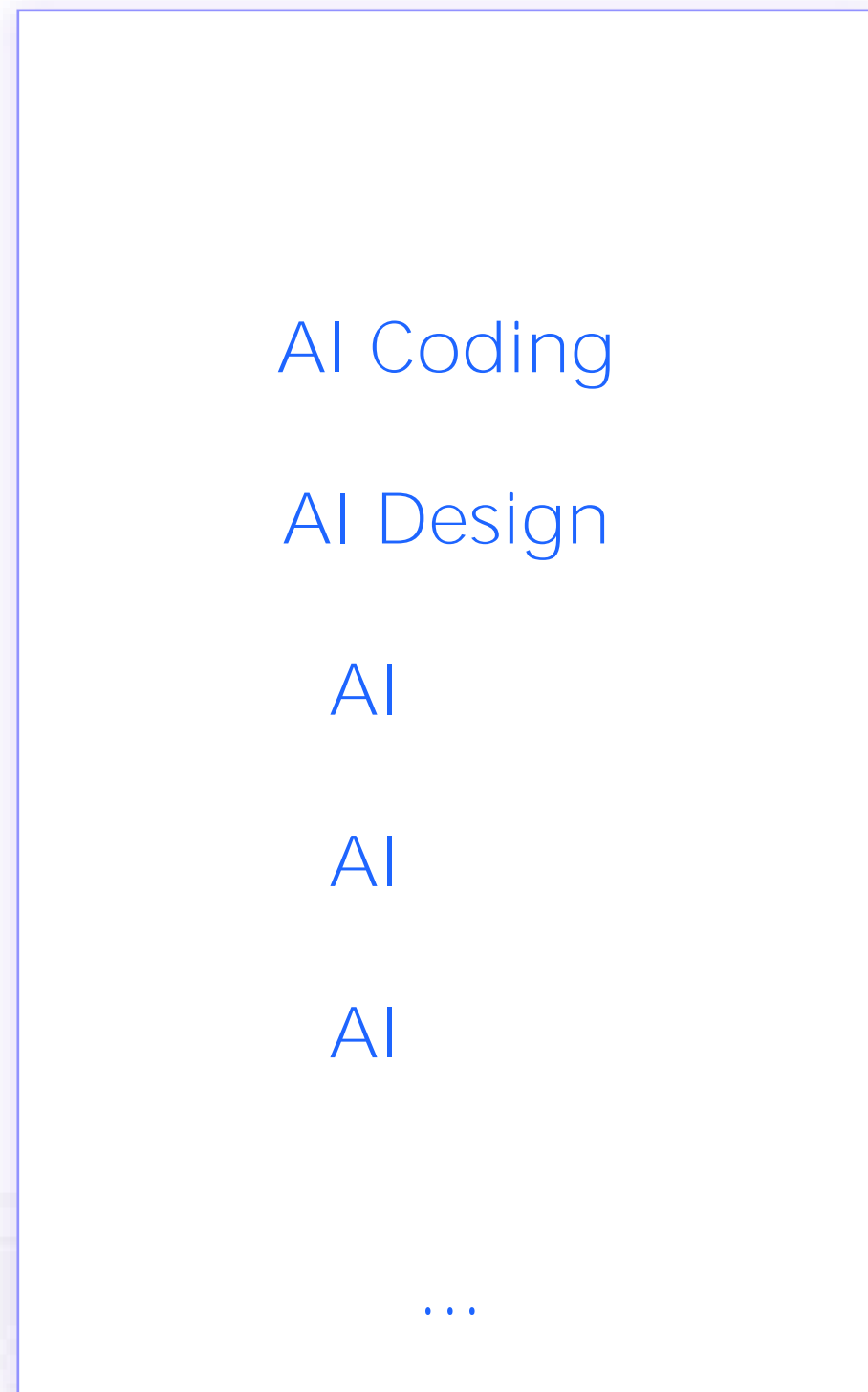
( )

---

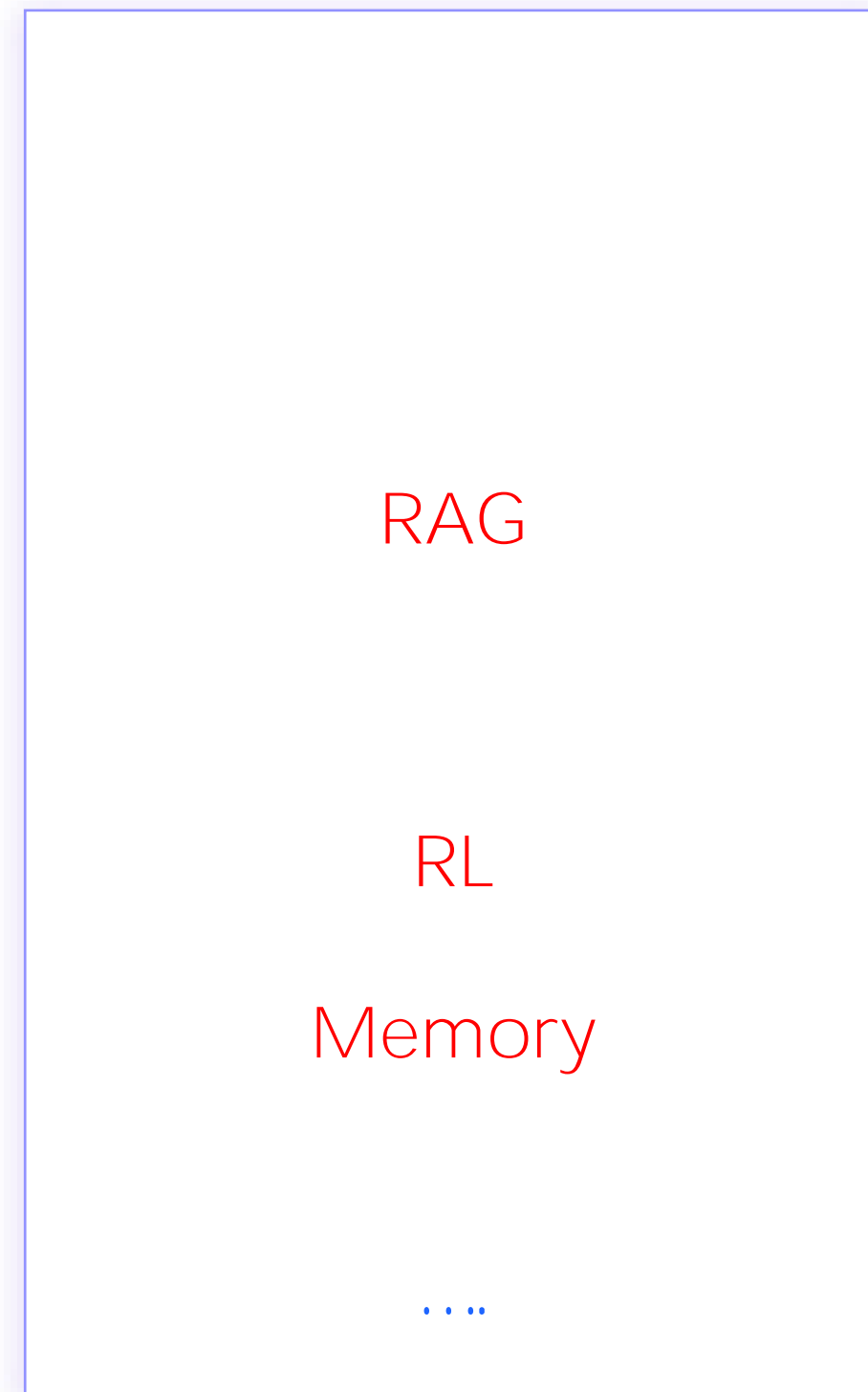


AI

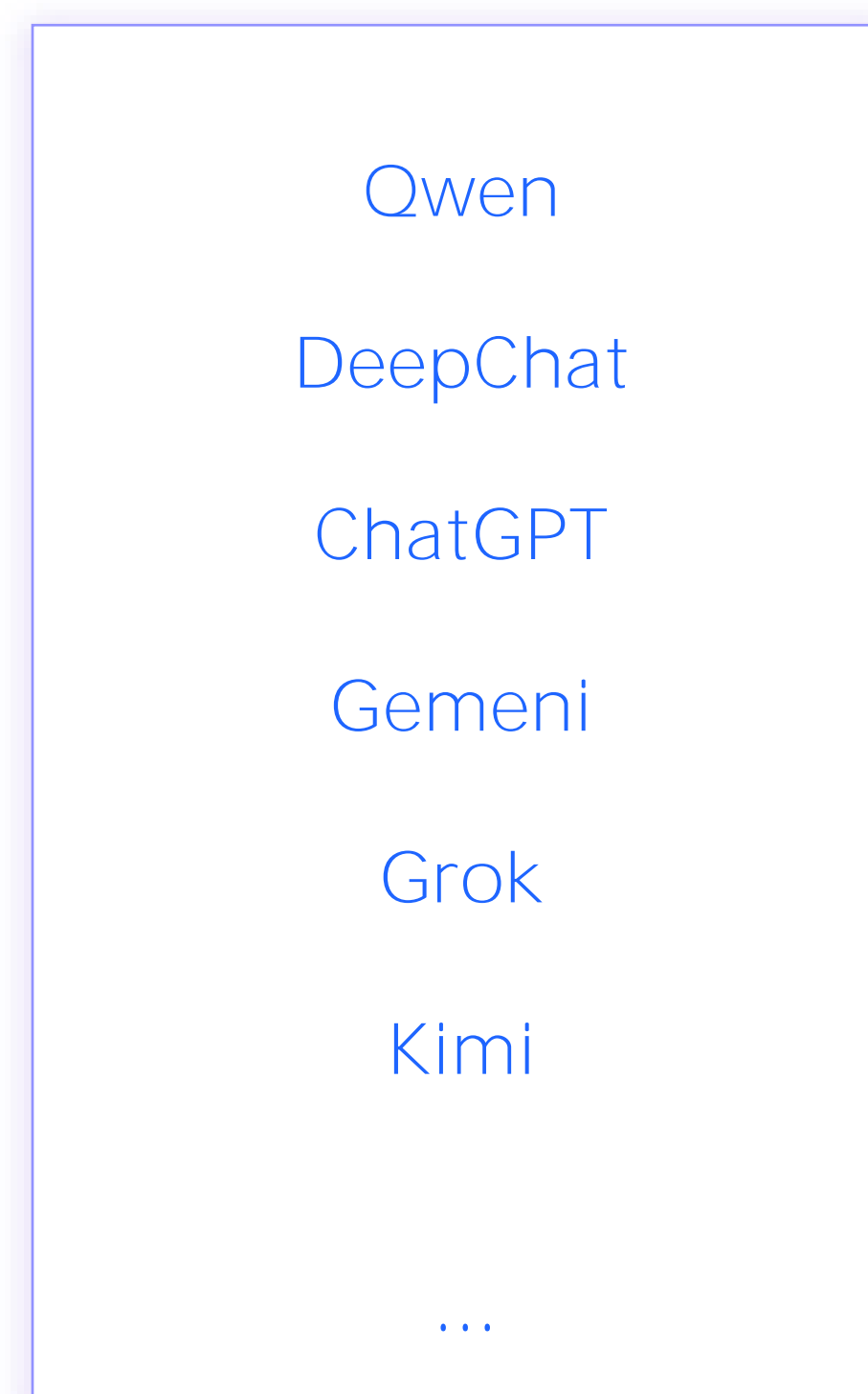
AI



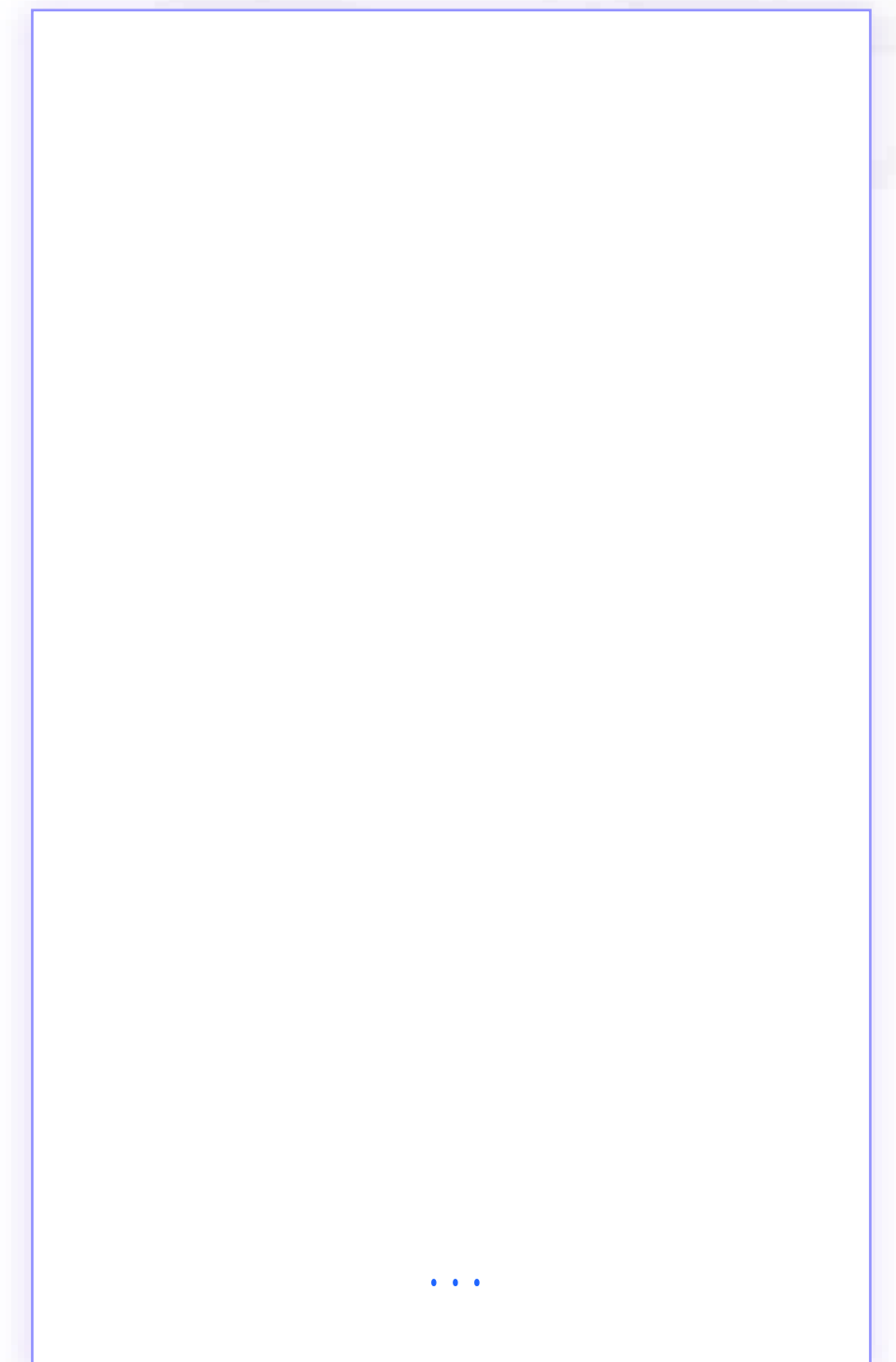
AI



AI



AI Infra



AI

4

\*



&

->

LangChain  
LlamaIndex  
Spring AI Alibaba...

2022 -2023

Function Calling  
MCP  
(Dify n8n...)

2023 -2024

Mem0  
Jina.ai

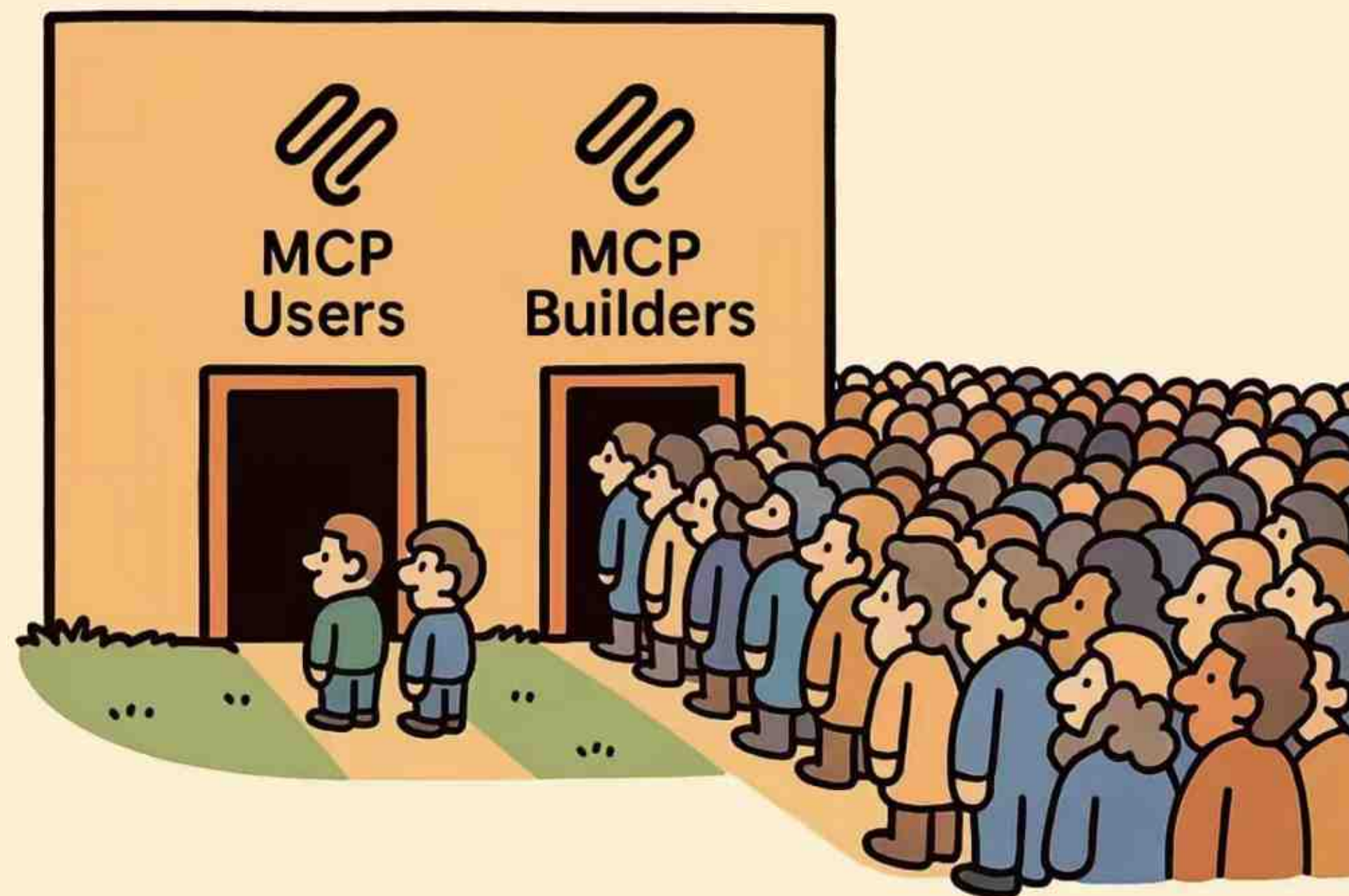
2024 -2025

OpenAI Kit  
Claude skills

2025 10 ~

\*

# Model Context Protocol



MCP is probably the only piece of tech that has more builders than users.

MCP

4

\*



&

->

LangChain  
LlamaIndex  
Spring AI Alibaba...

2022 -2023

Function Calling  
MCP  
(Dify n8n...)

2023 -2024

Mem0  
Jina.ai

2024 -2025

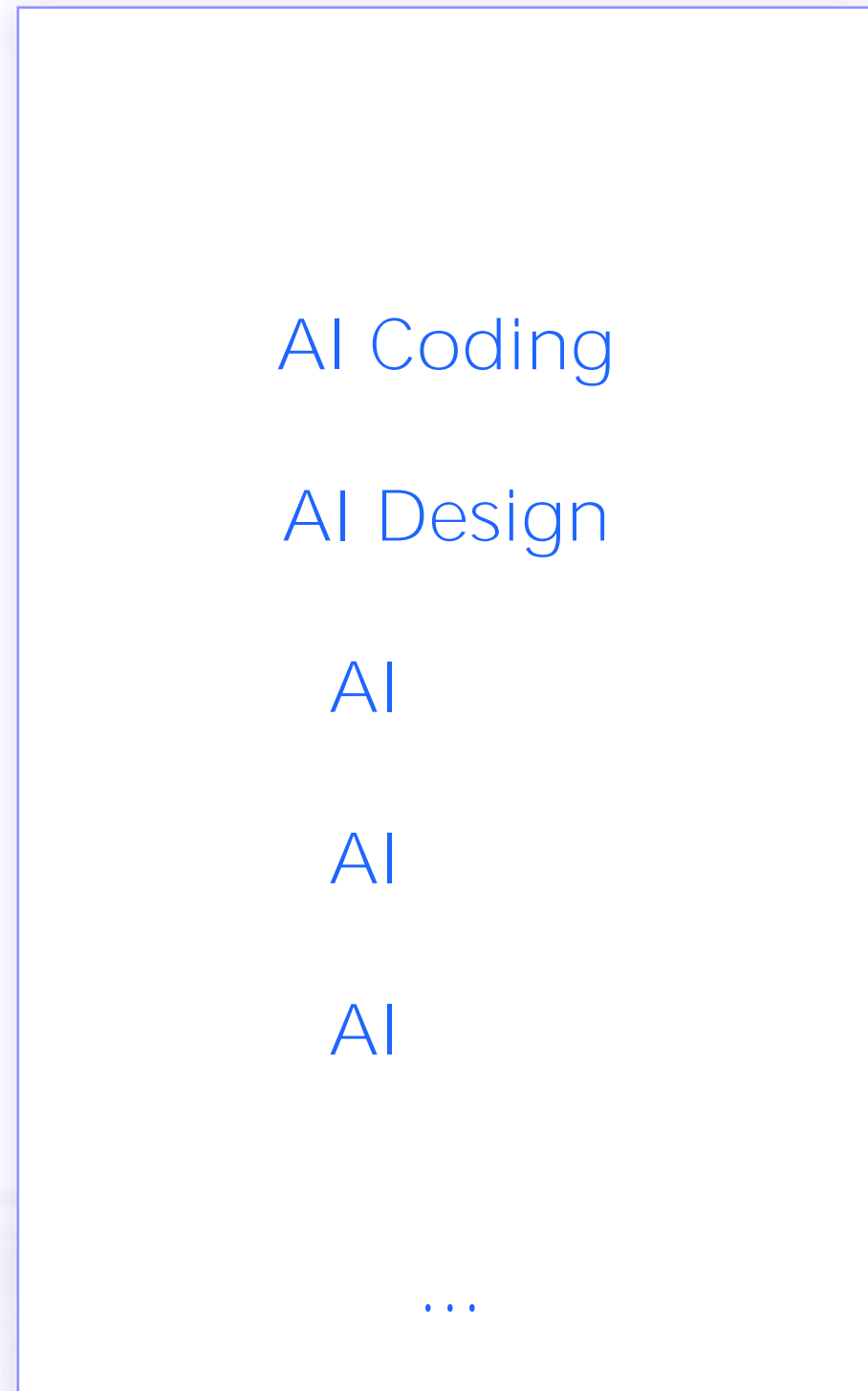
OpenAI Kit  
Claude skills

2025 10 ~

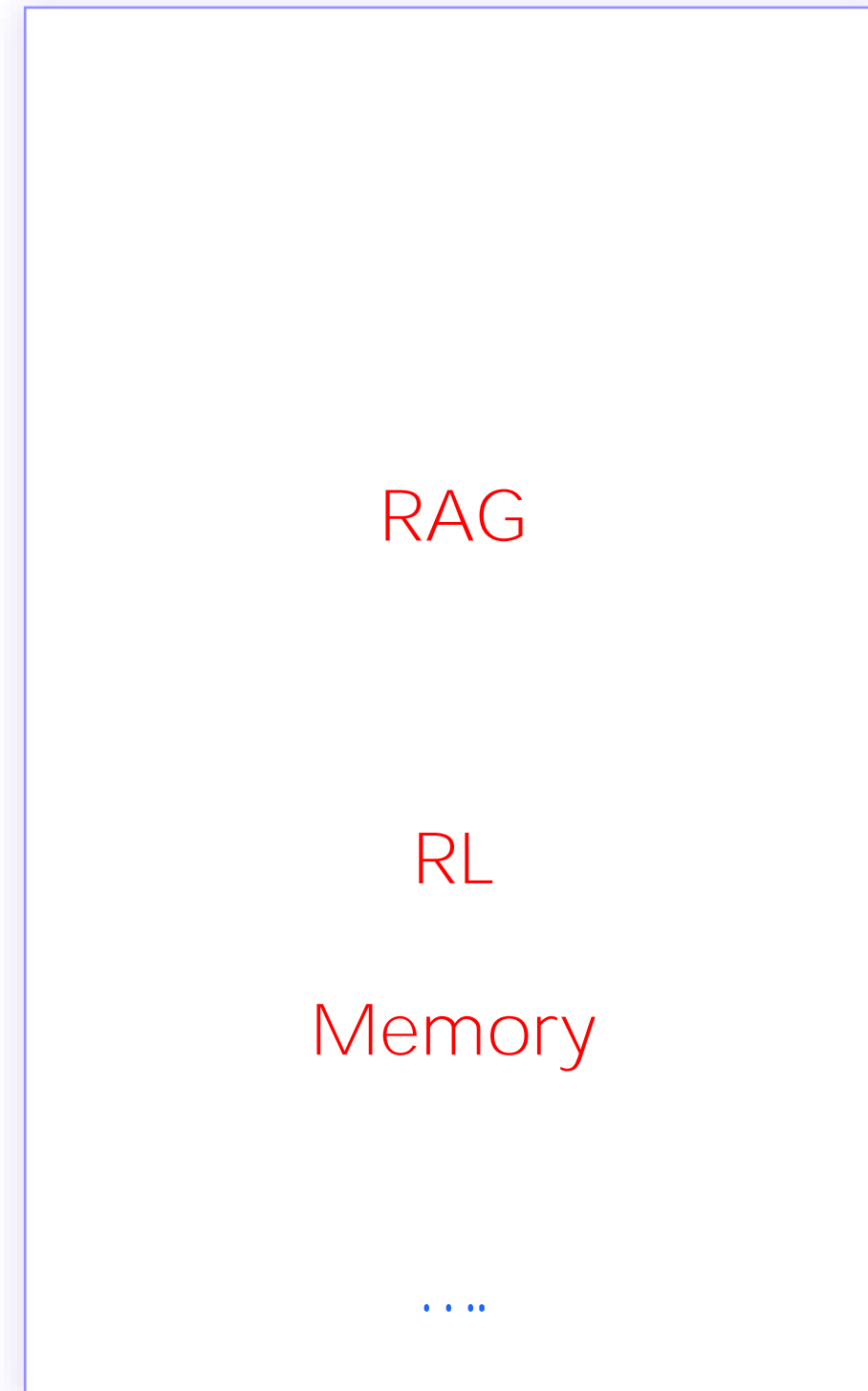
\*

AI

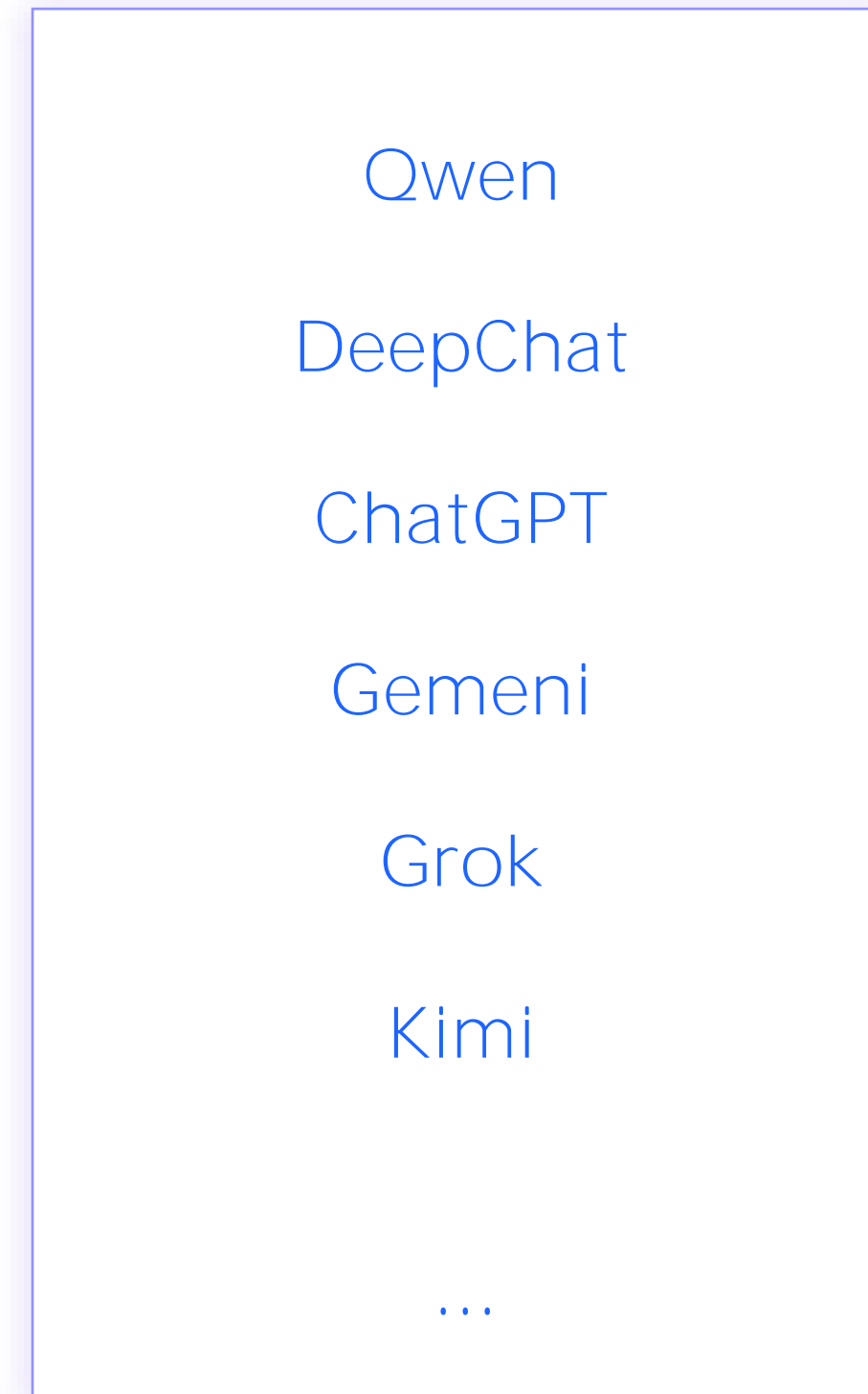
AI



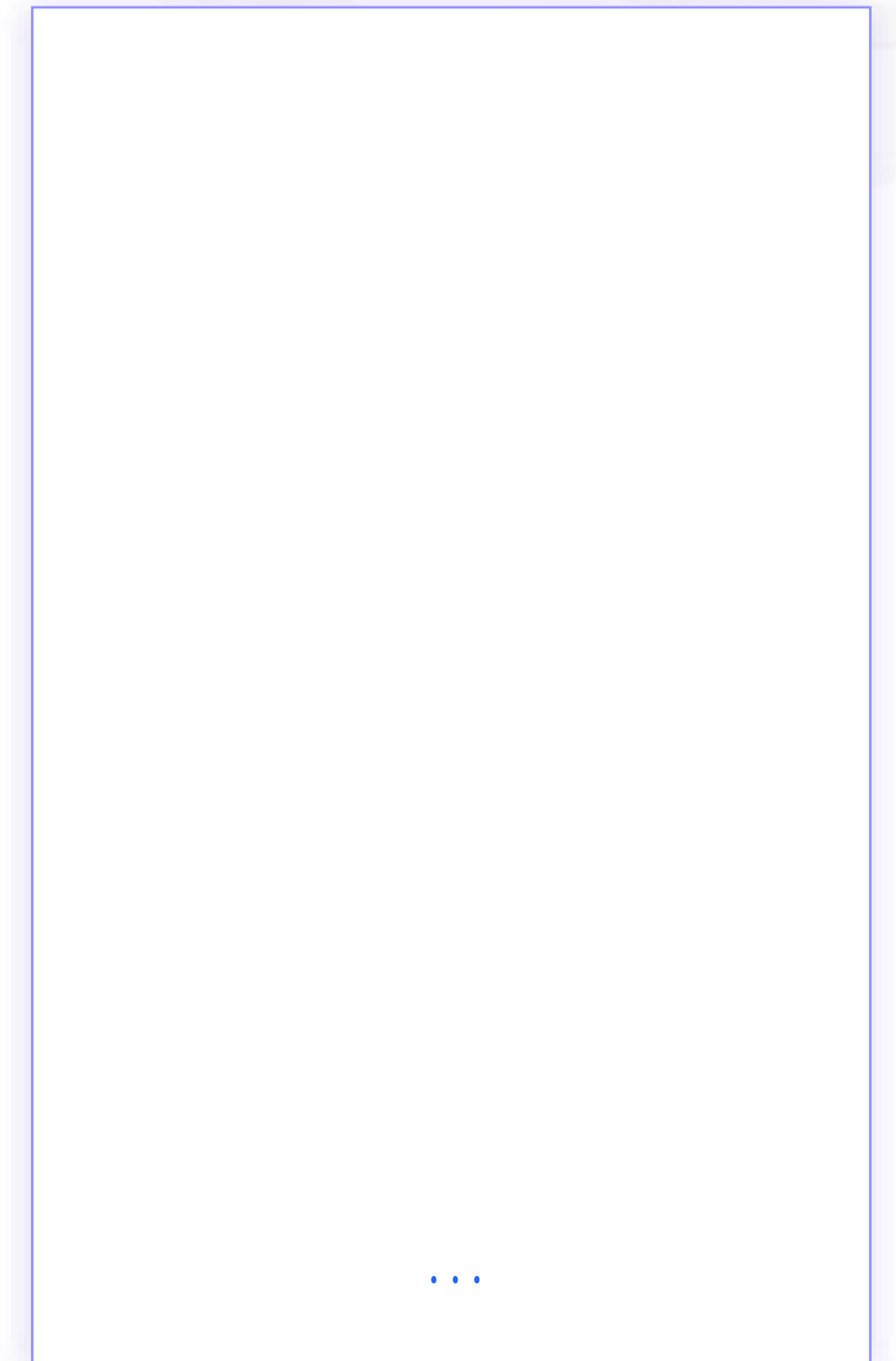
AI



AI



AI Infra



AI

AI



1

2

3

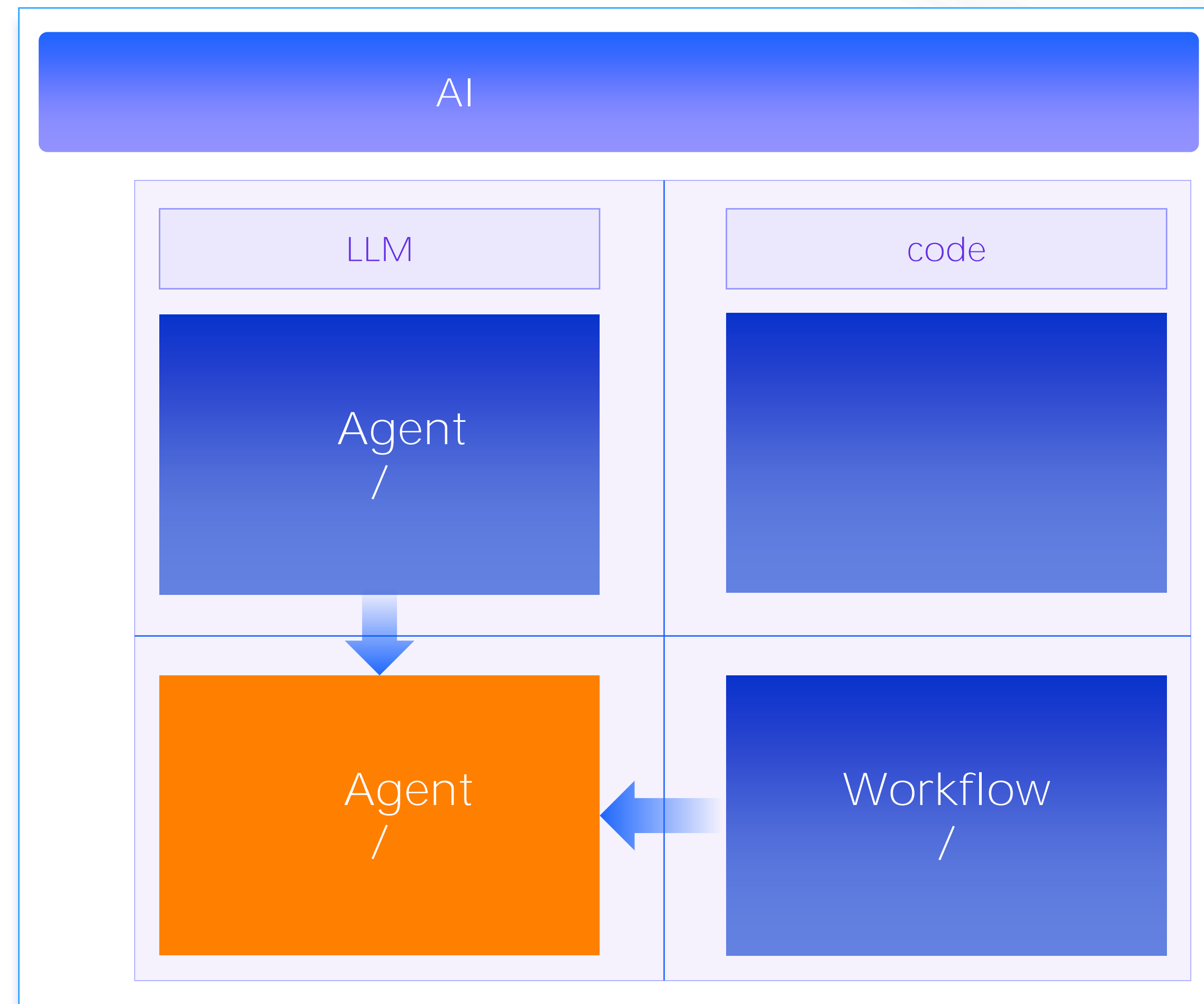
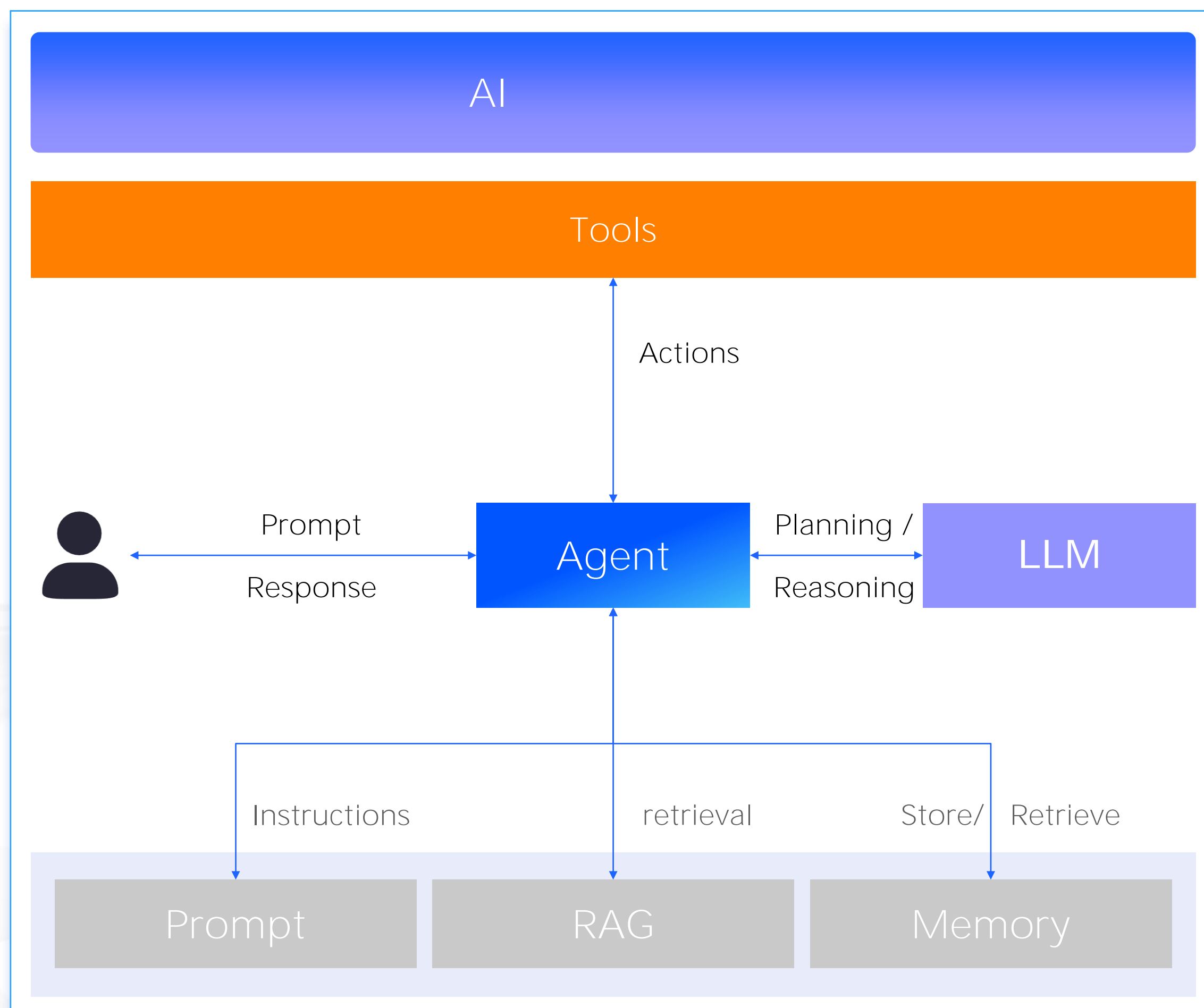
4

+

+

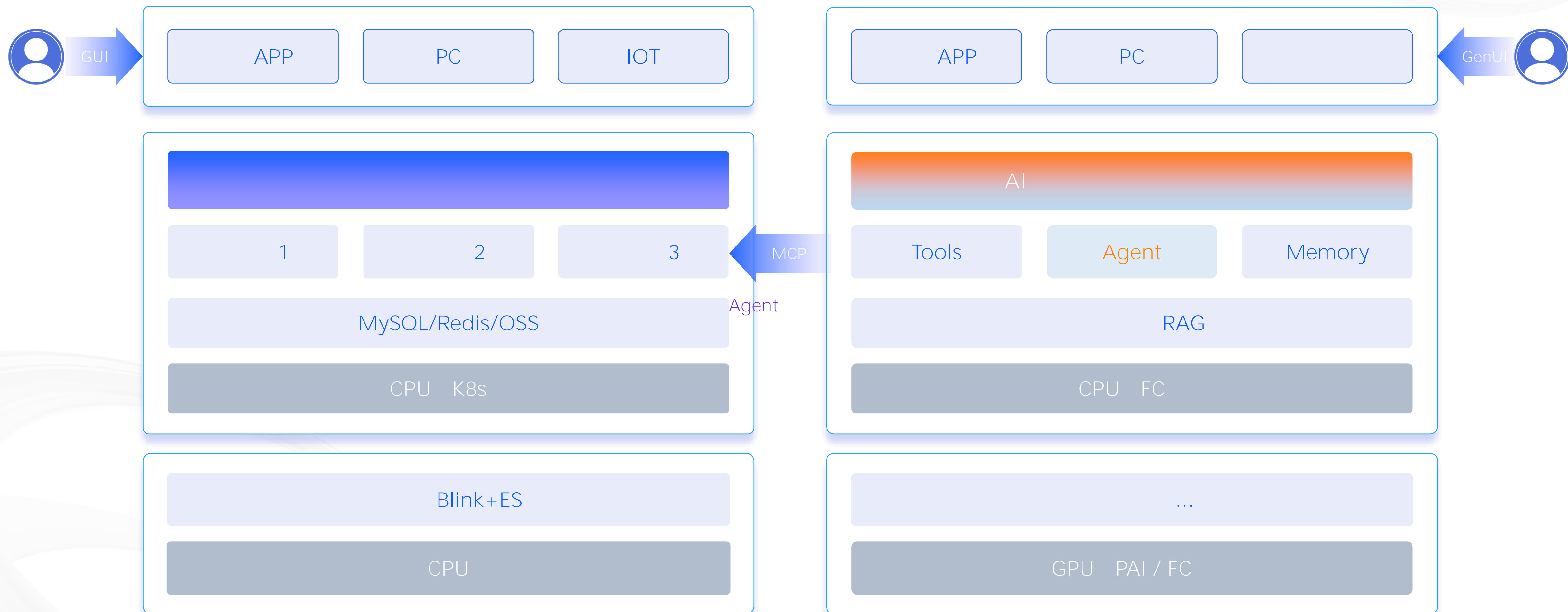
# AI

## Agent

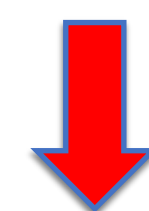
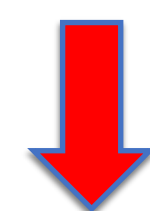


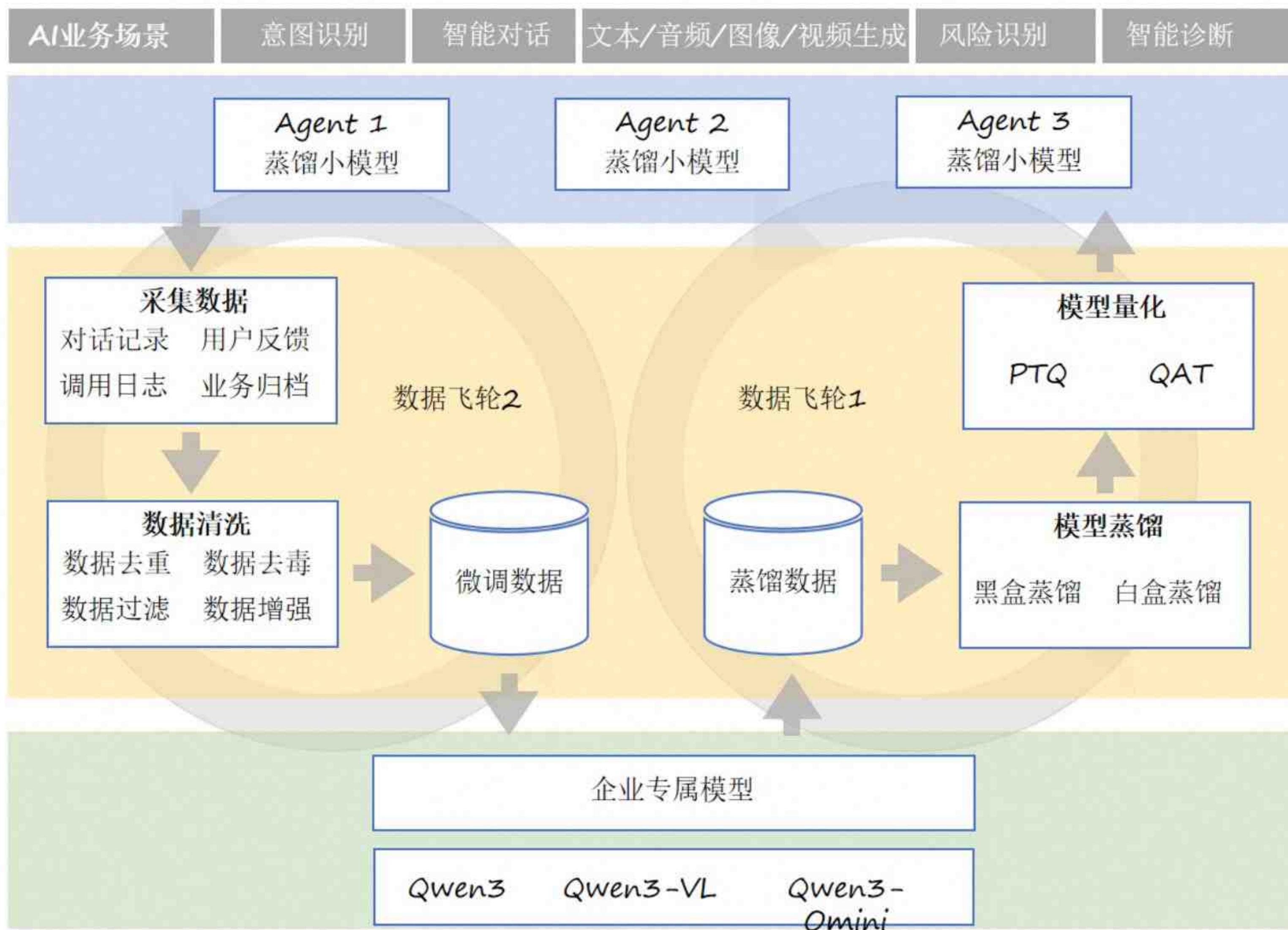
# AI

+AI



AI





**数据飞轮1**，每个垂类场景形成小飞轮，准备好企业的专属数据，用大尺寸SOTA模型作为教师模型来生成高质量蒸馏数据集，用来微调小尺寸模型，用于支持垂类Agent

算力需求：少量算力卡完成7B~32B模型的SFT

**数据飞轮2**，多个Agent在实际业务场景下的生产数据，比如应用日志经过打标，成为新的训练数据，反过来增强大尺寸模型，让教师模型逐渐具有企业/行业属性，成为企业专属模型底座

算力需求：大量算力卡完成较大规模CPT/SFT/RL，实现大飞轮

# AI



AI  
11



15



20 +



40+

THANKS

# Serverless AI

AI

A Unified Hosting Platform for Legacy and AI Applications

Serverless



# CONTENT

01

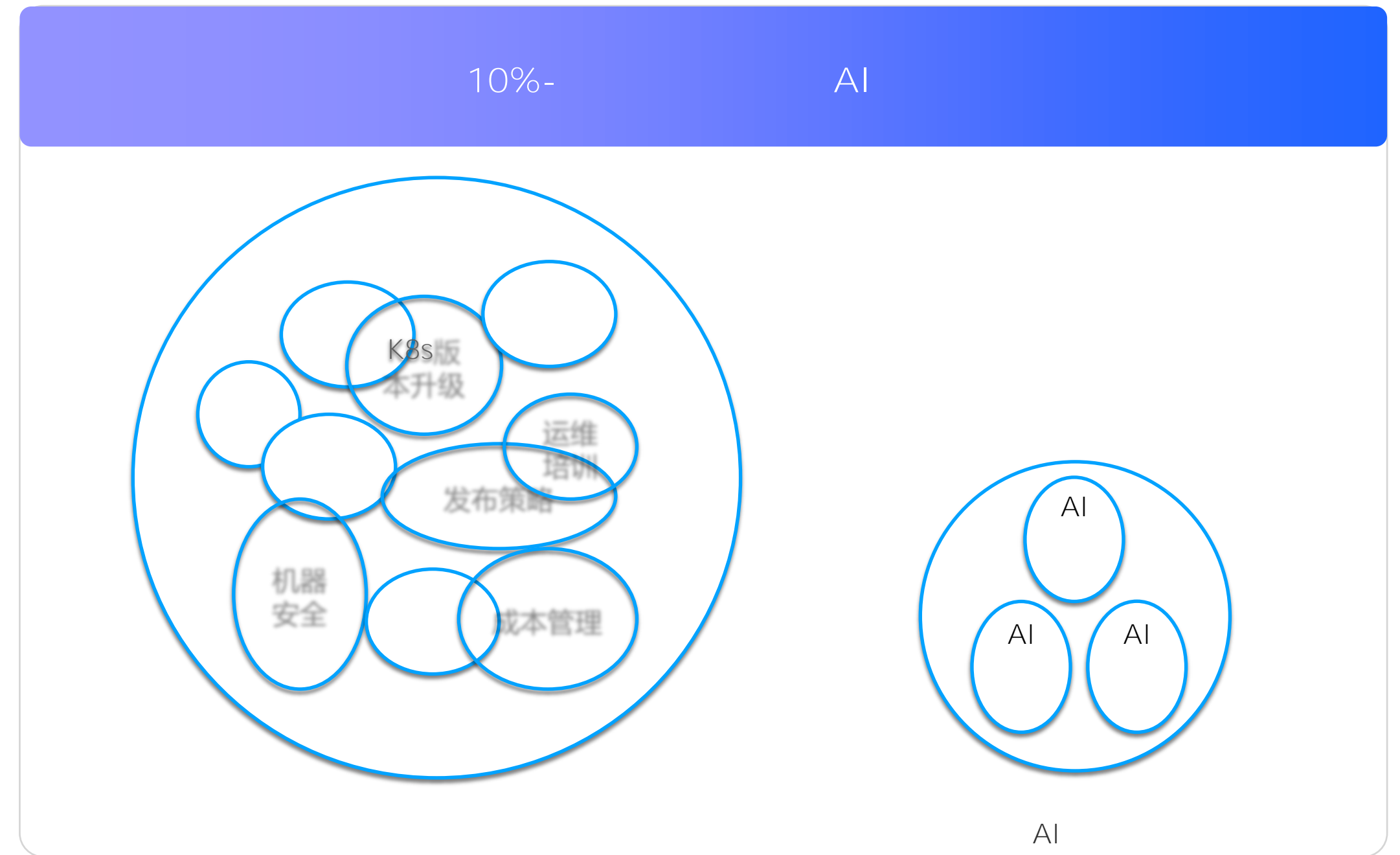
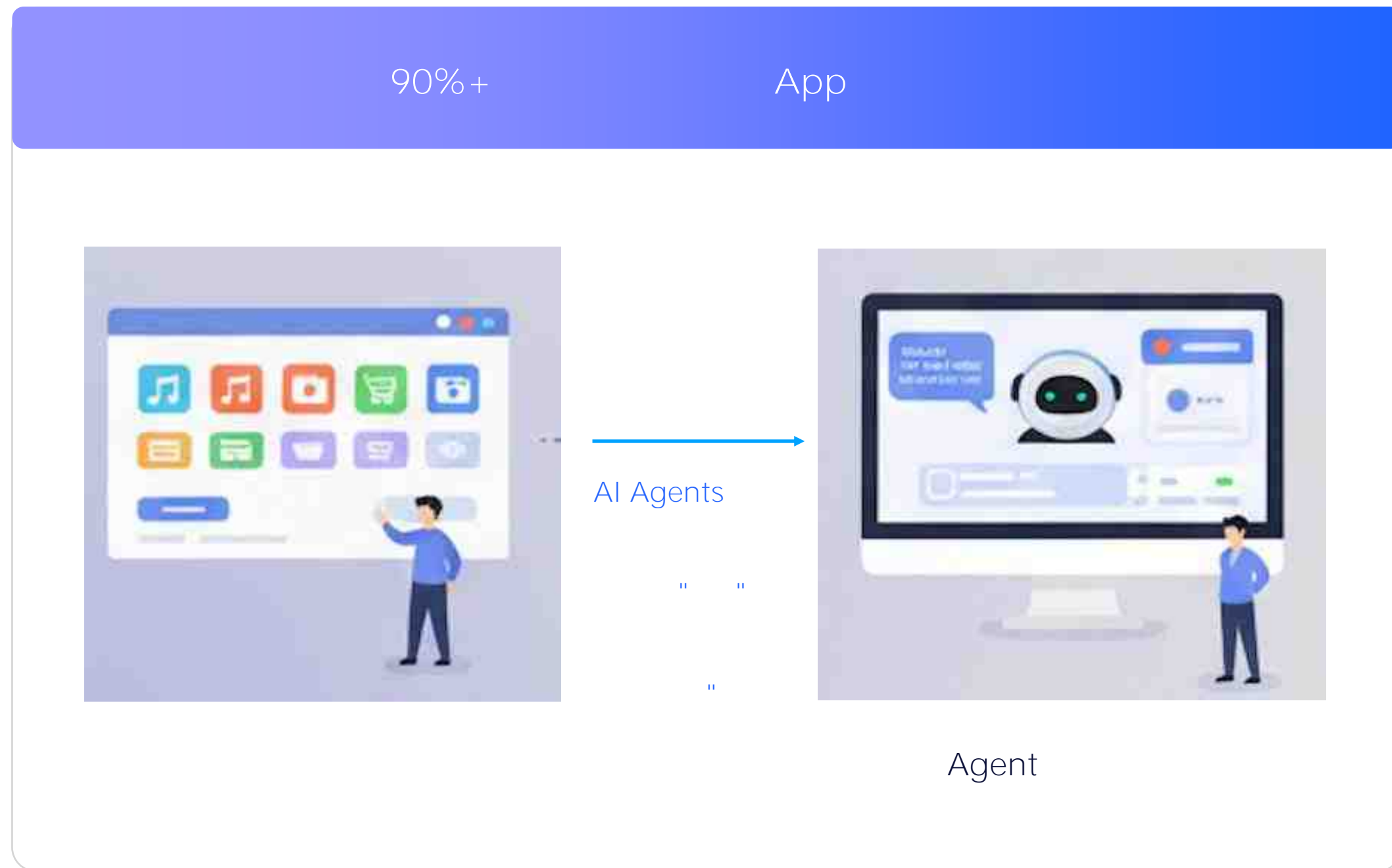
AI

02

AI

# AI

## AI Adoption Surges, but Investment Structures Remain Imbalanced



1990 / 2020 /  
/ AI Agents/ /

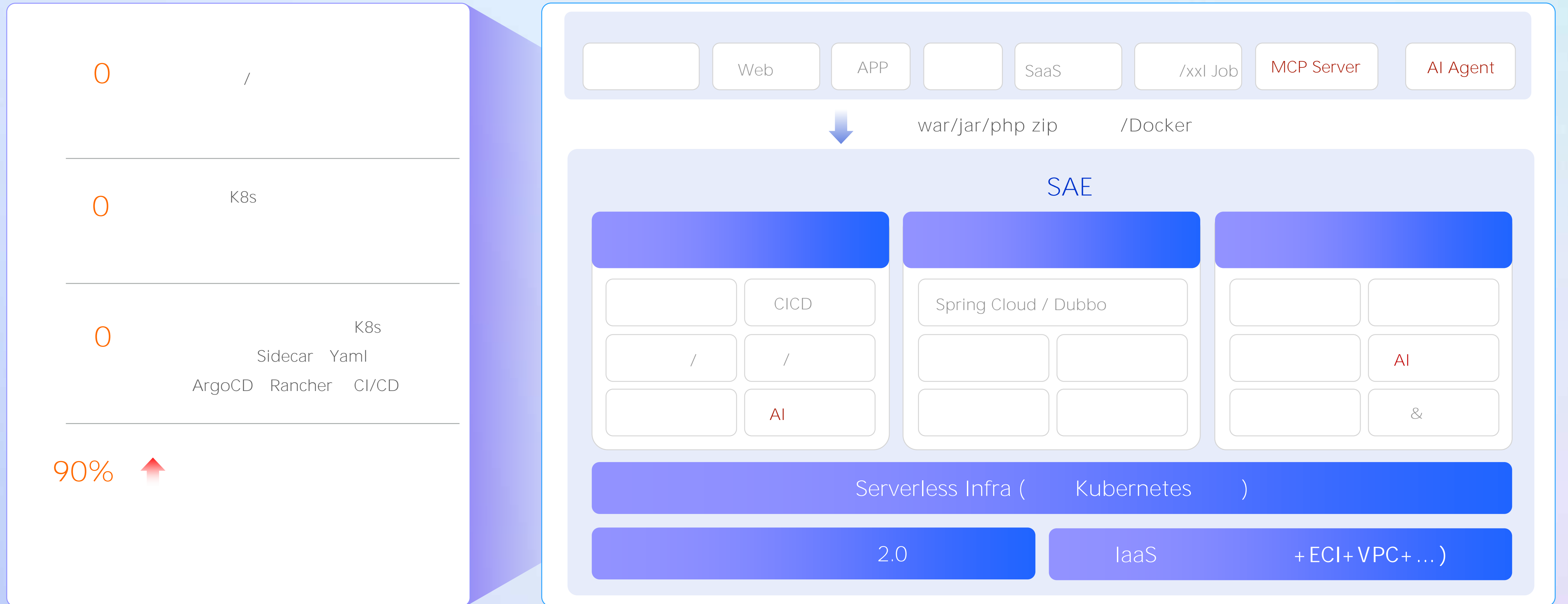
• / / ROI •  
• / / • KPI

01

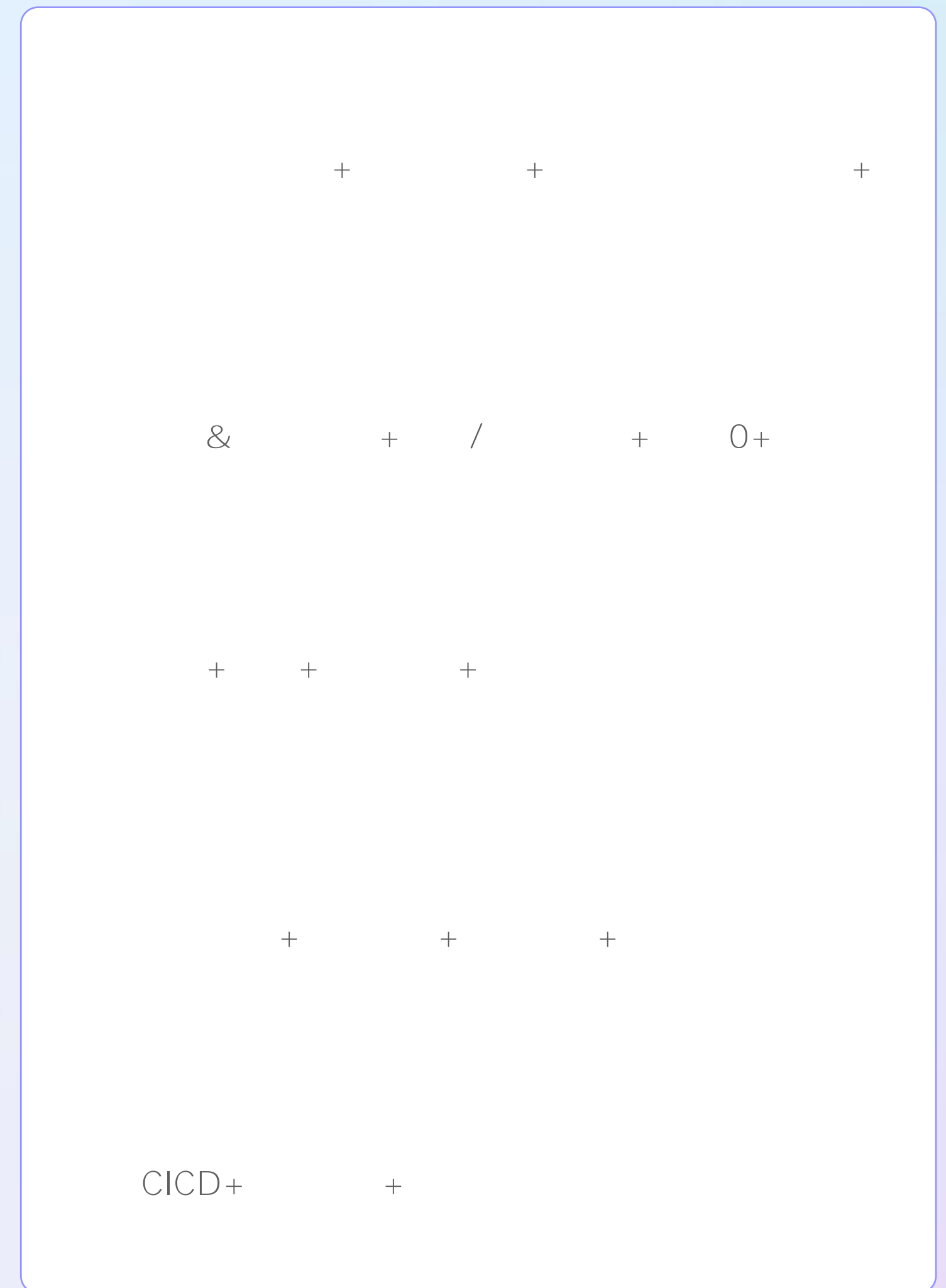
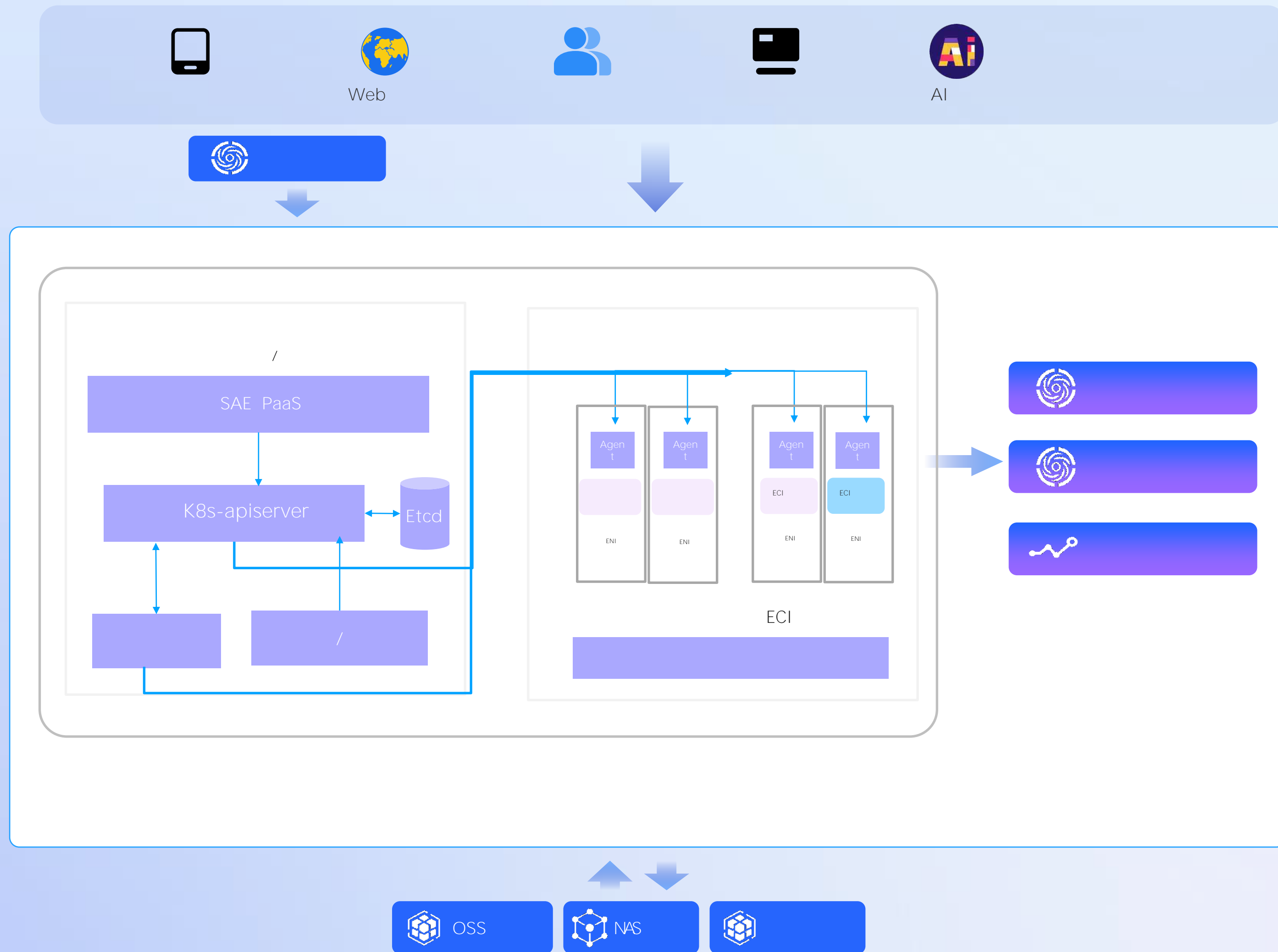
--Serverless App Engine SAE

### Simple: One-stop Container Hosting, Simpler Operations

SAE



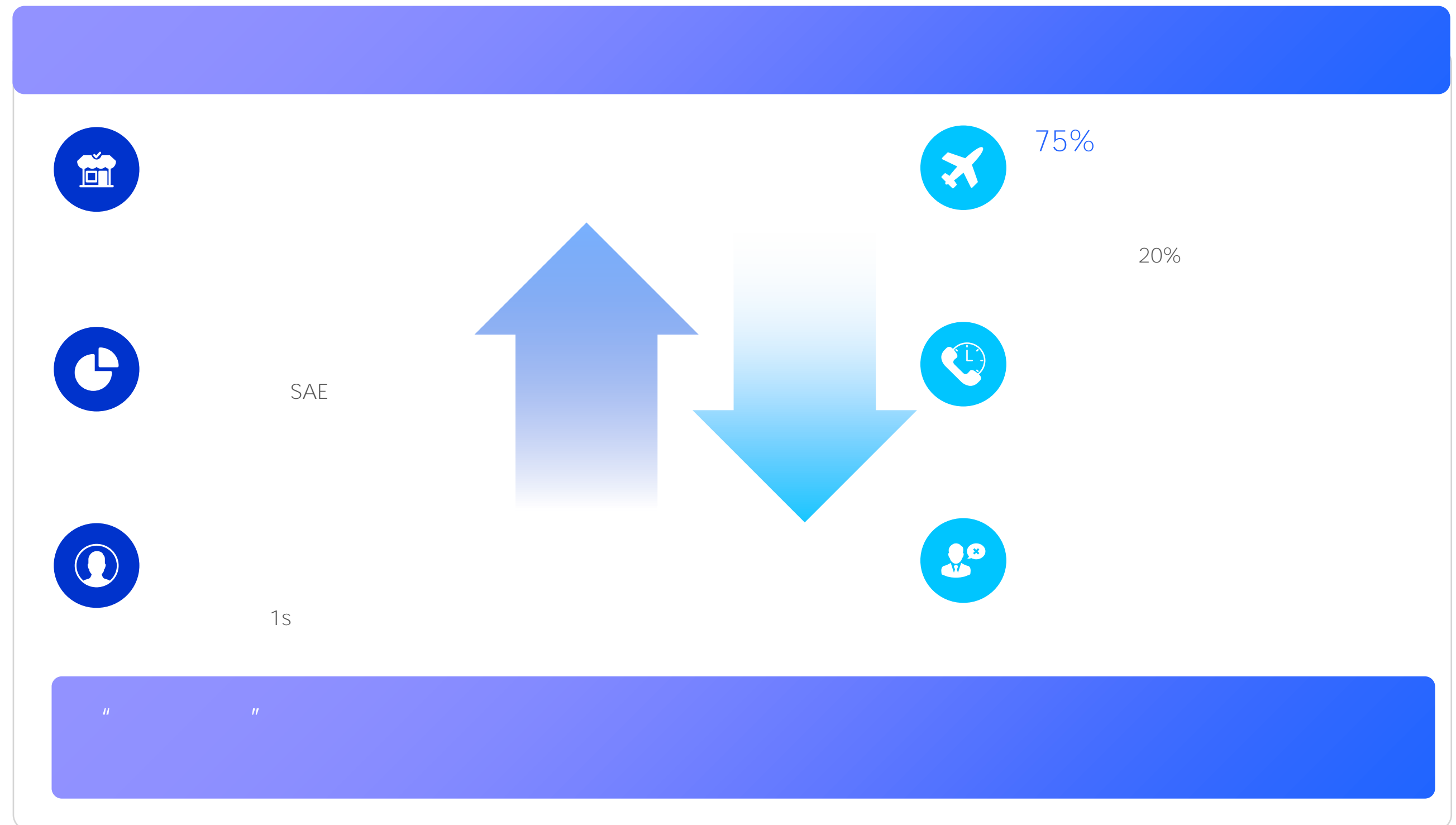
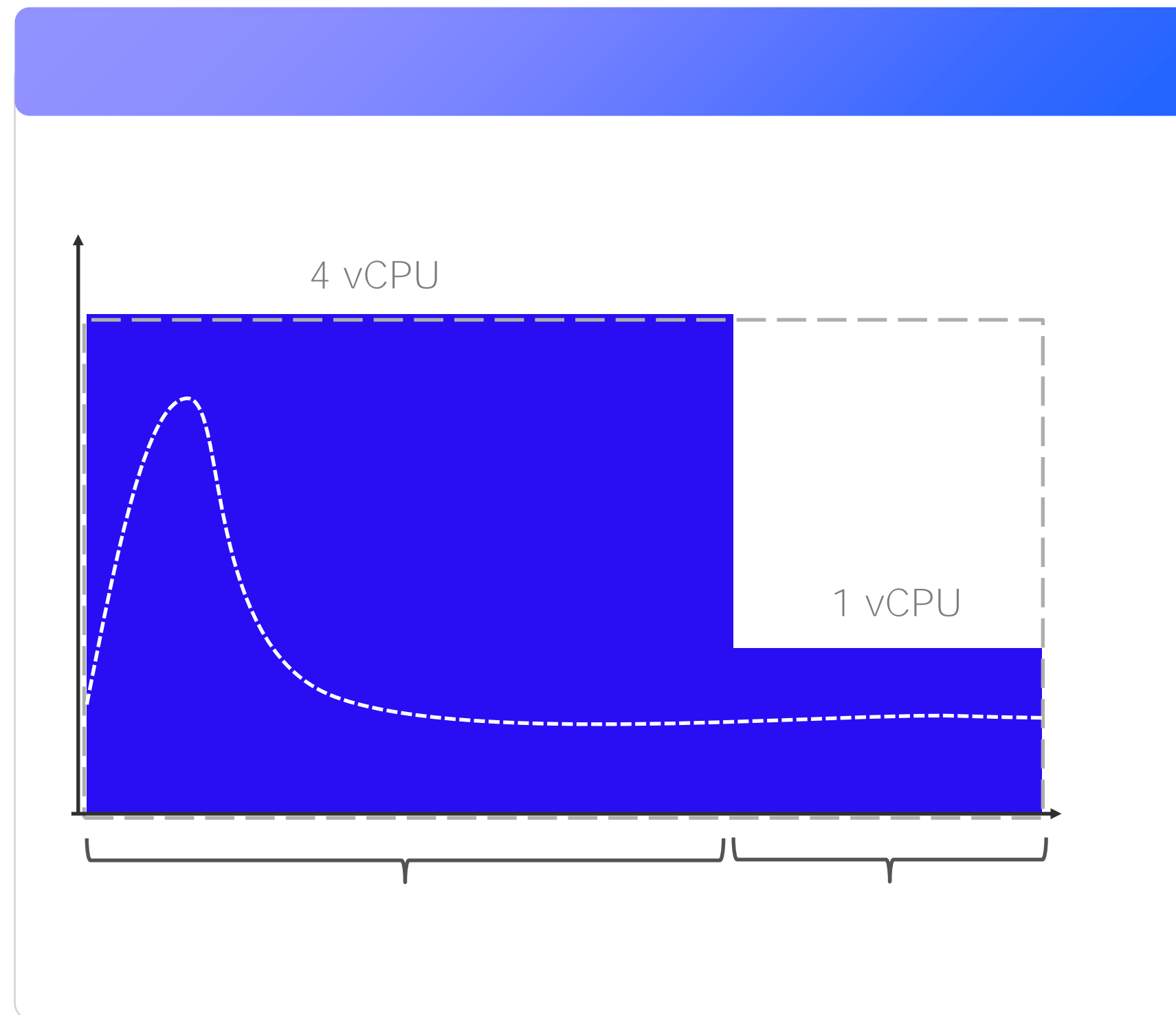
# Stable: End-to-End Architecture and Security for Uninterrupted Online Operations



+ +

### Cost-Efficient: Maximizing Resource Utilization through Pay-as-You-Go, Zero Redundancy, and Overselling Benefits

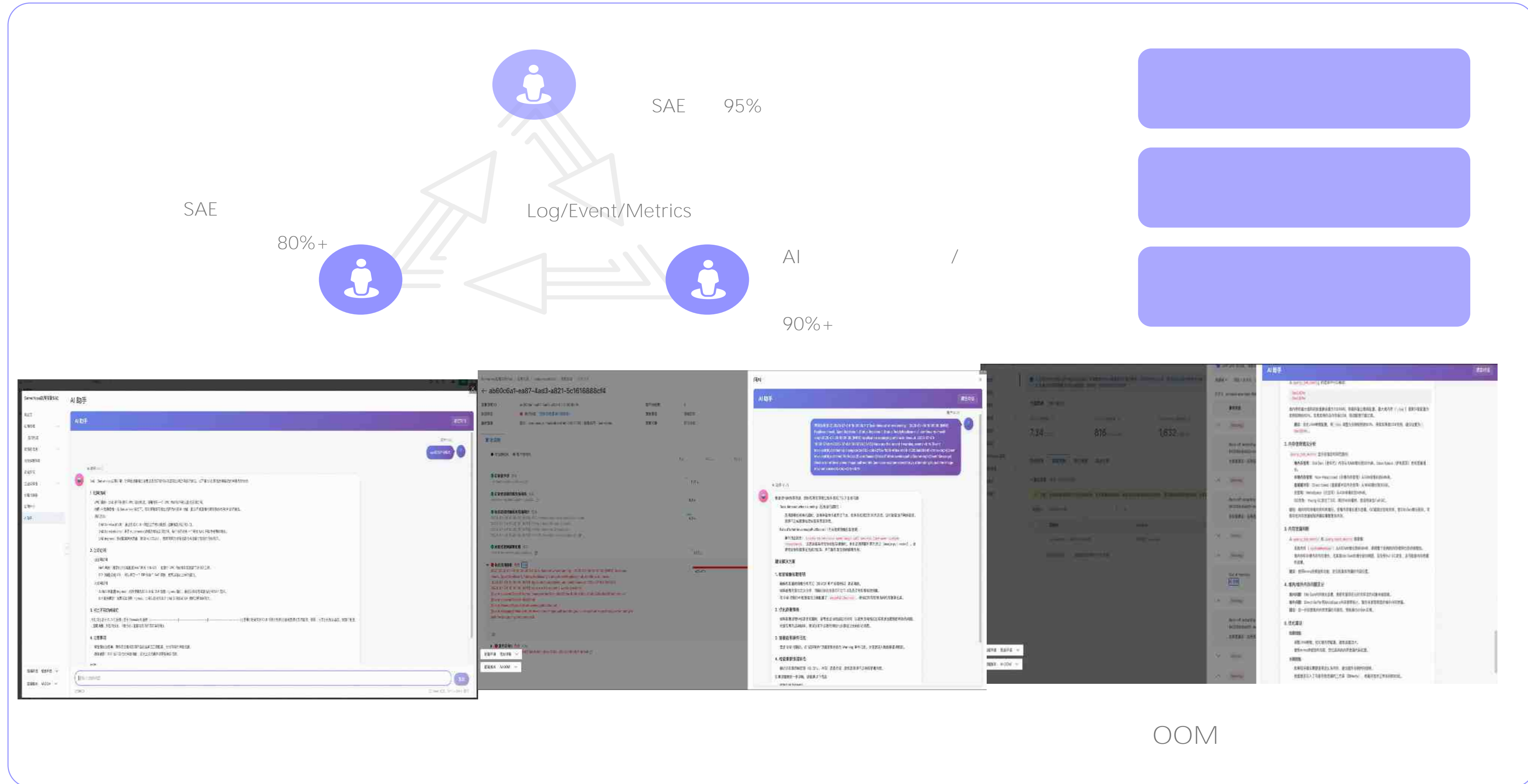
75%



# SAE

## AI Intelligent Assistant Solutions and Capabilities

AI



# 02 AI

-- AgentRun

# Agent

Bridging the Agent Productivity Gap

AI



&

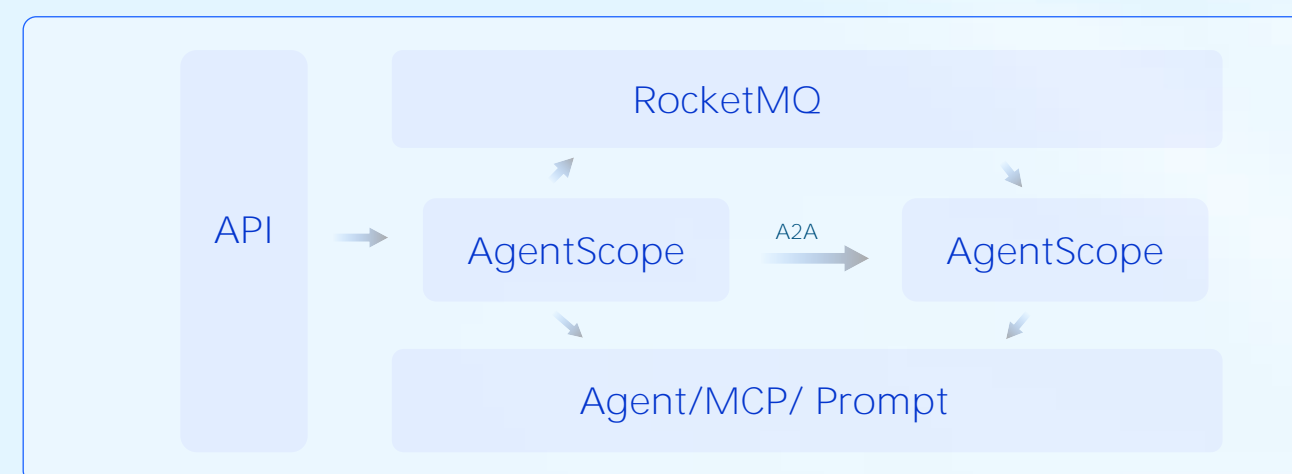
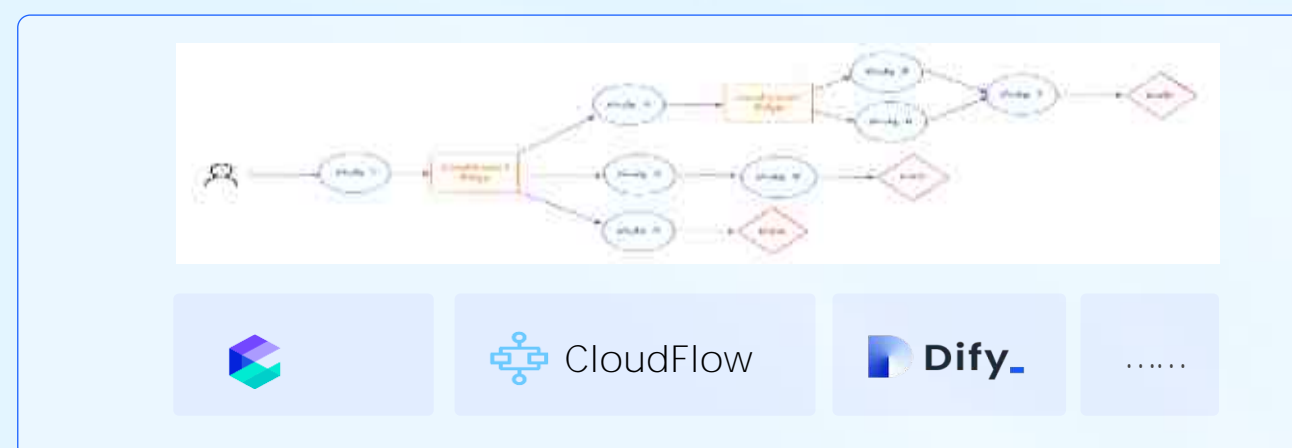
Muti-Agent

# AgentRun Agentic AI

FunctionCompute AgentRun: Agentic AI Application Infrastructure

Agentic AI

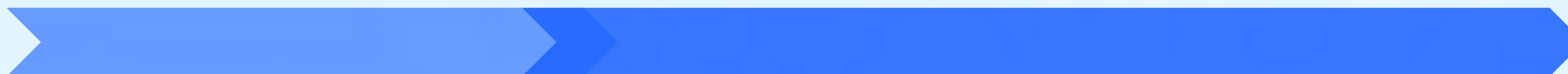
AI



# AgentRun Agentic AI

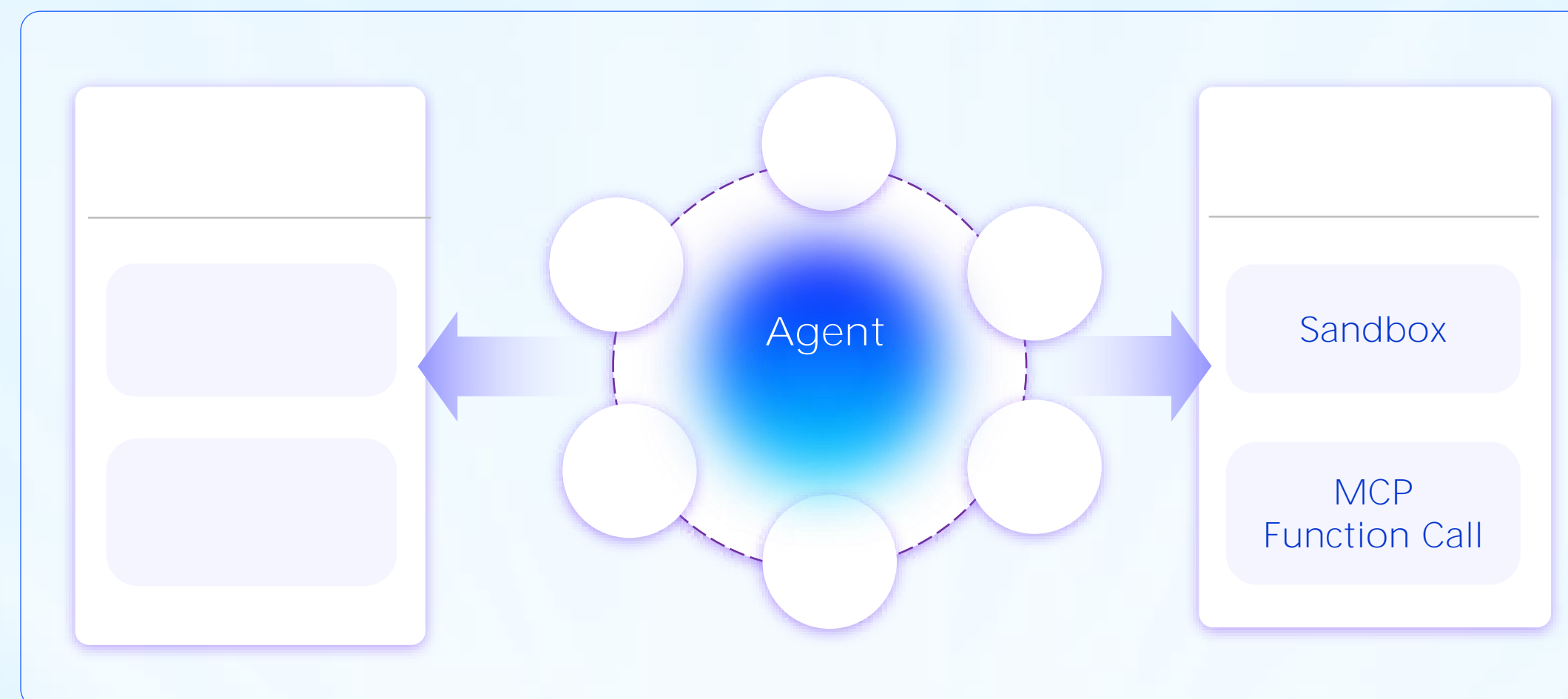
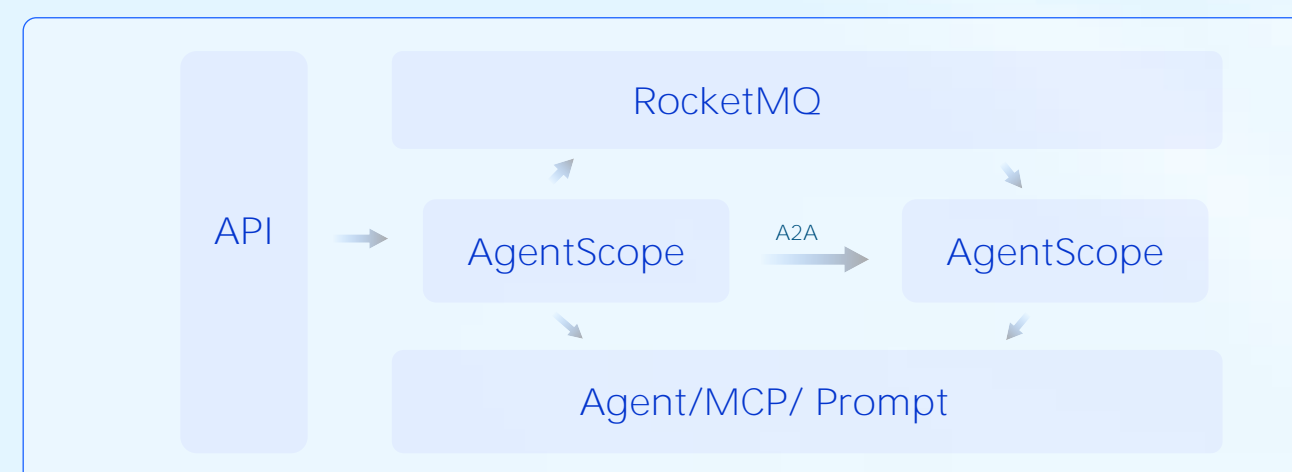
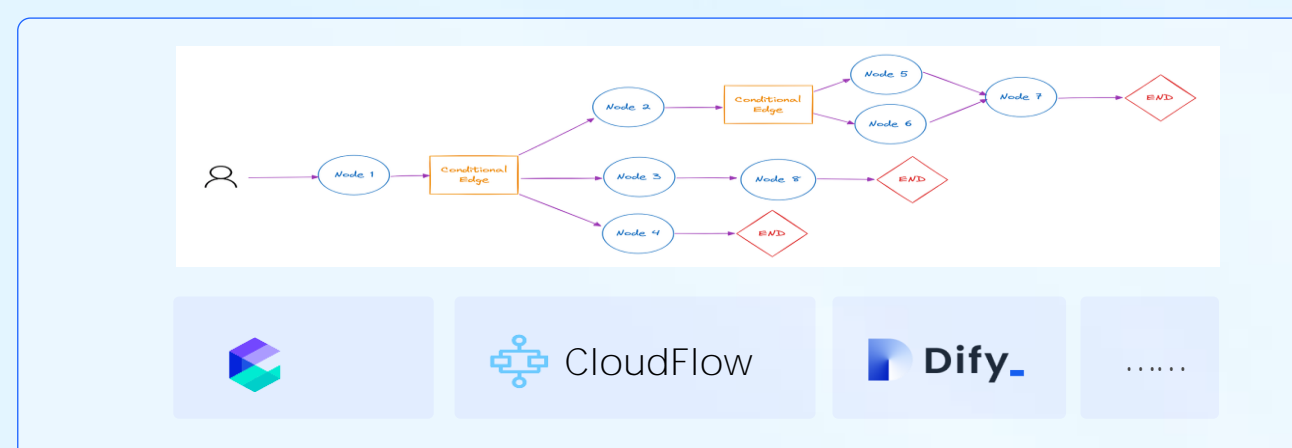
FunctionCompute AgentRun: Agentic AI Application Infrastructure

Agentic AI



AI

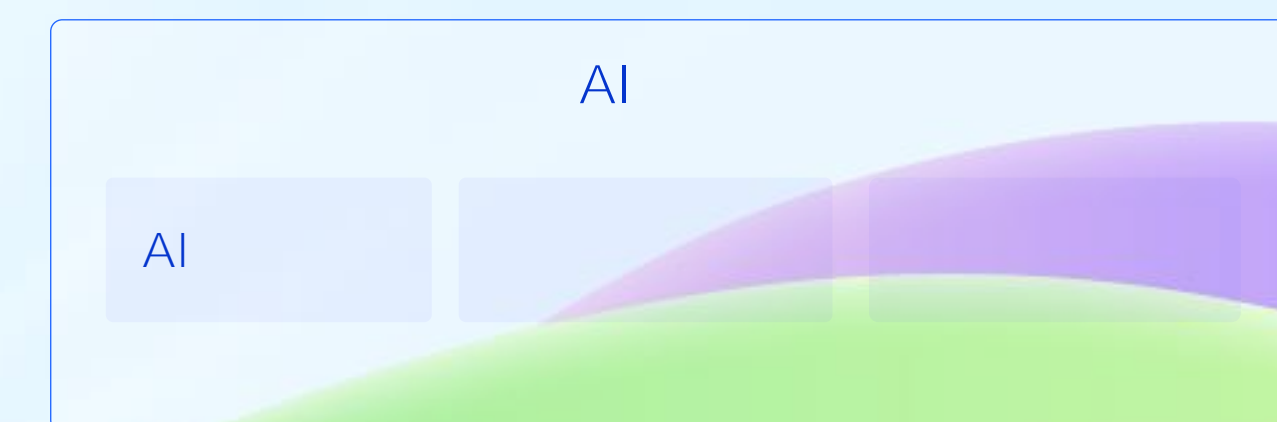
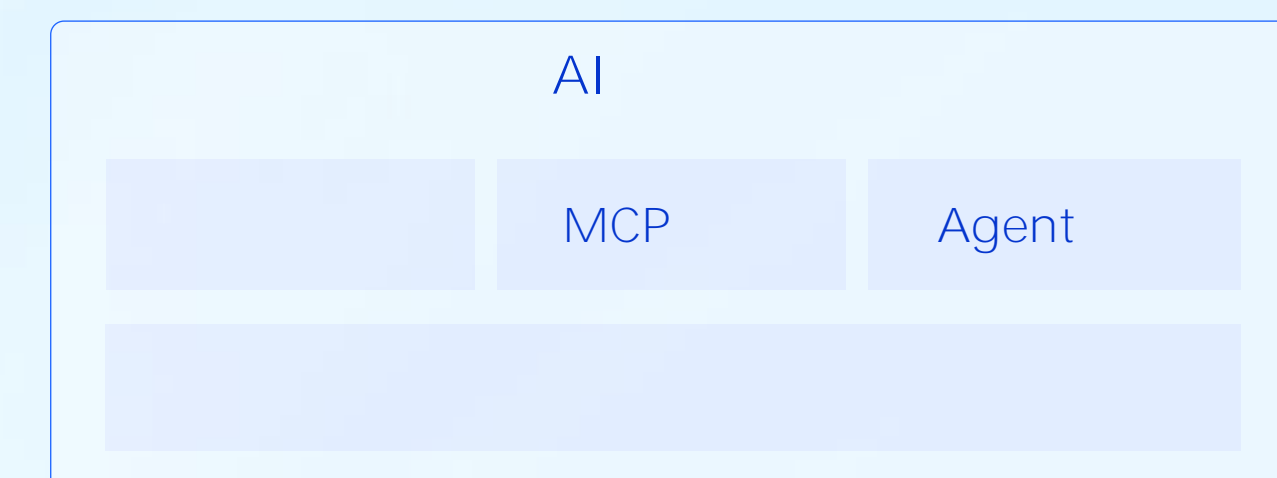
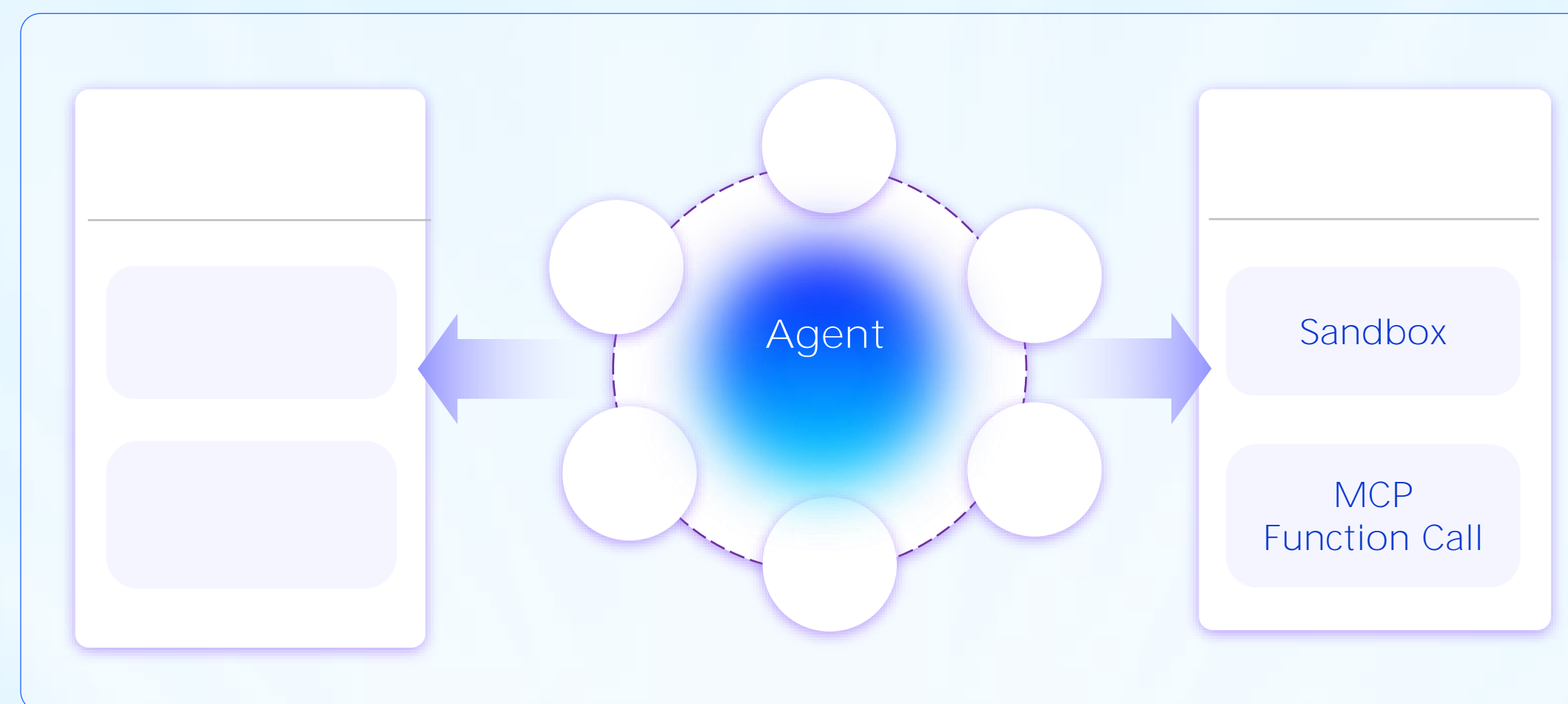
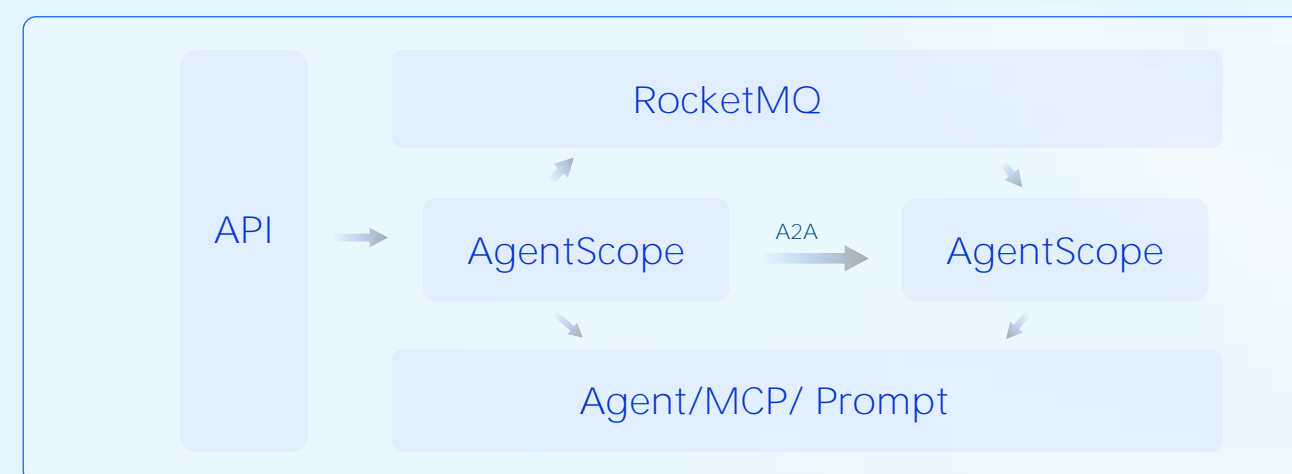
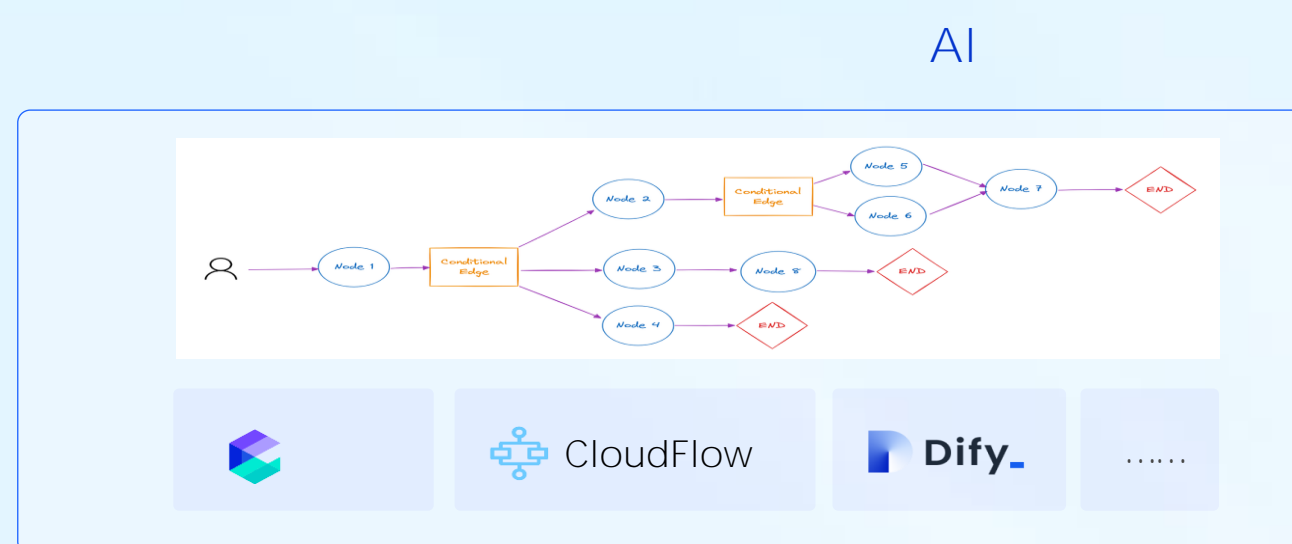
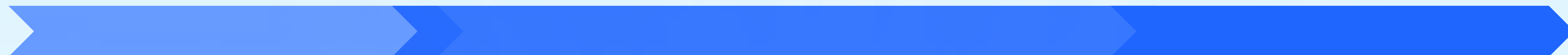
Serverless AI



# AgentRun Agentic AI

FunctionCompute AgentRun: Agentic AI Application Infrastructure

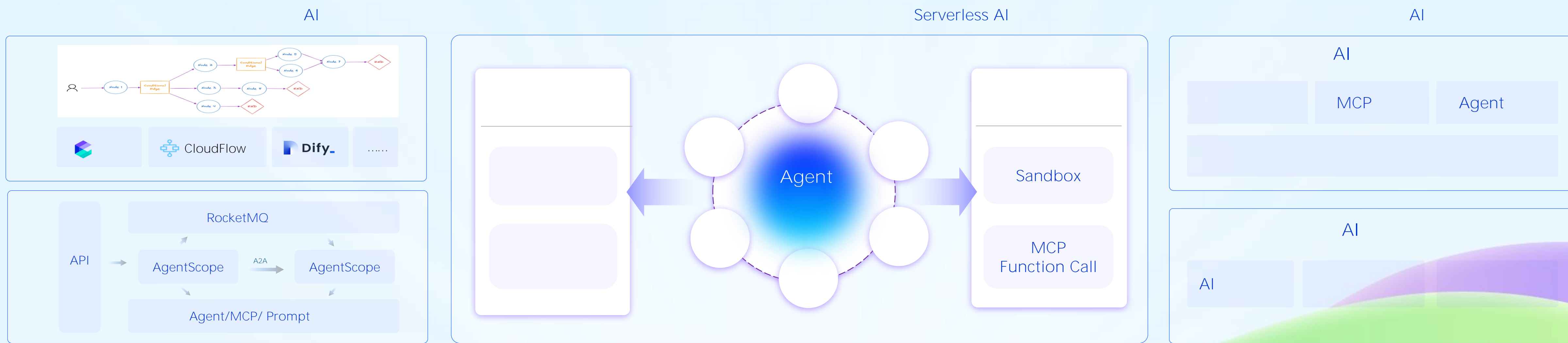
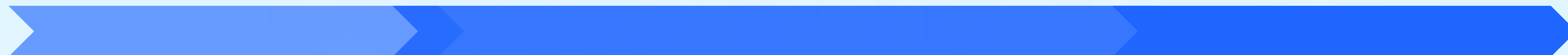
## Agentic AI



# AgentRun Agentic AI

FunctionCompute AgentRun: Agentic AI Application Infrastructure

## Agentic AI



## AgentRun

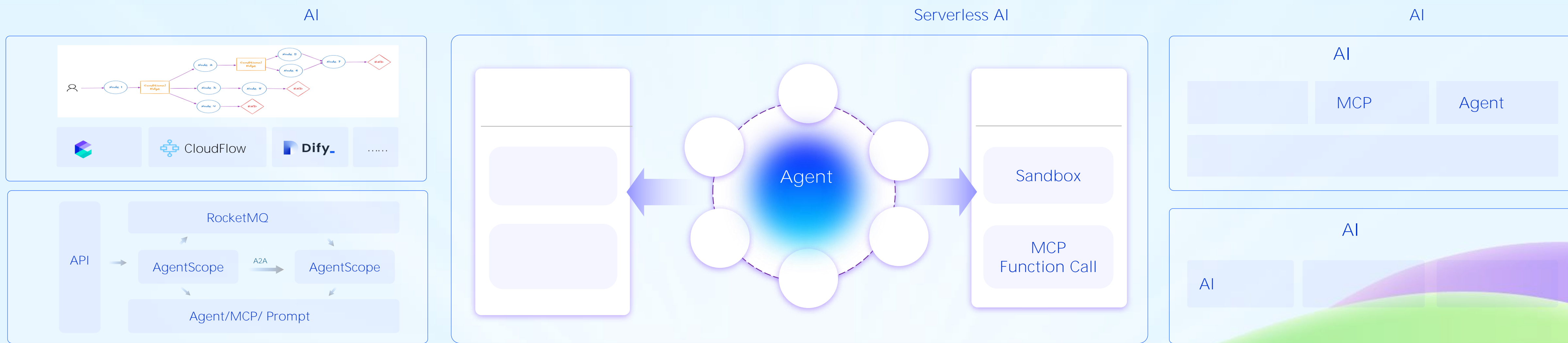
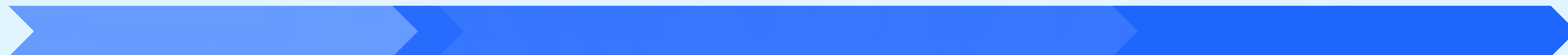
Agent

AI

# AgentRun Agentic AI

FunctionCompute AgentRun: Agentic AI Application Infrastructure

## Agentic AI



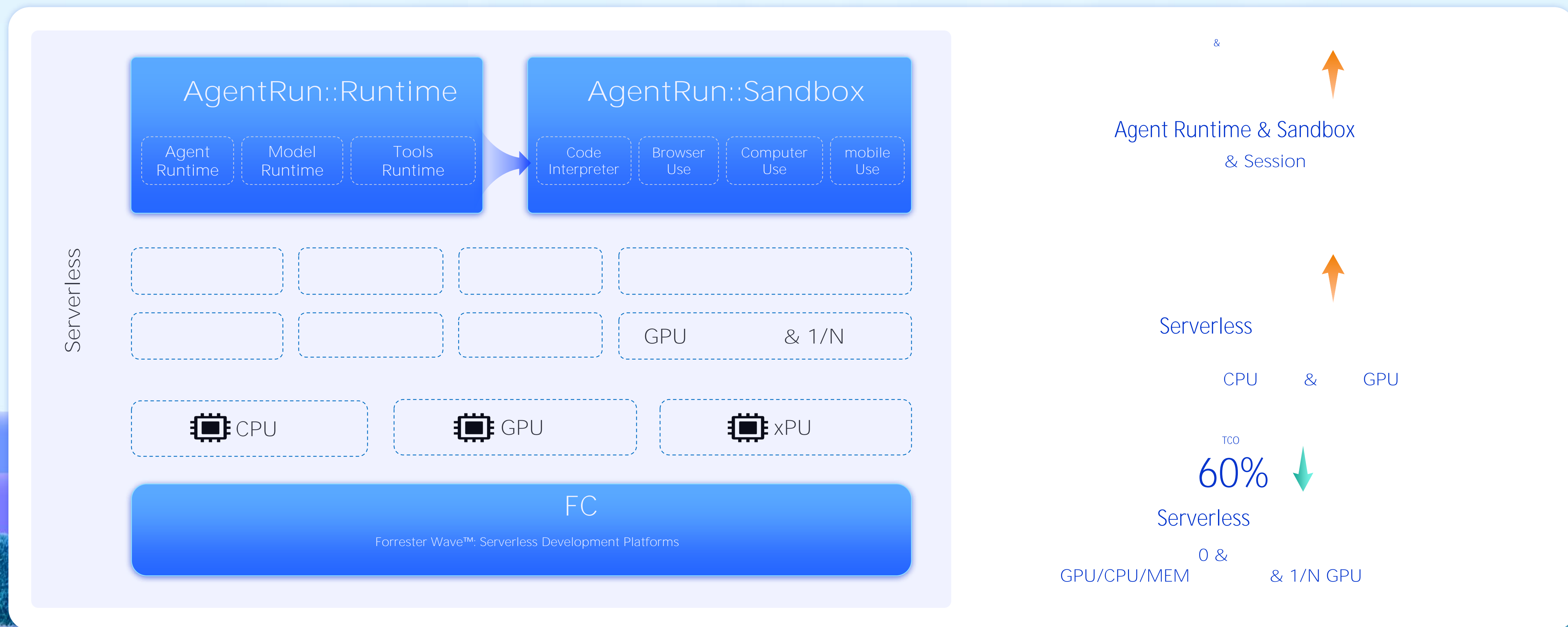
## AgentRun



# AgentRun &

AgentRun Agent Runtime & Sandbox

Serverless AI



# AgentRun


## AgentRun Model Runtime

FC

1/N




**FC**

 RT 80%

独家优势, 相比 HPA 抖动率明显下降

**FC**

 以 15% 的成本换取冷启动分钟级->毫秒级

**1/N**

 93.75%

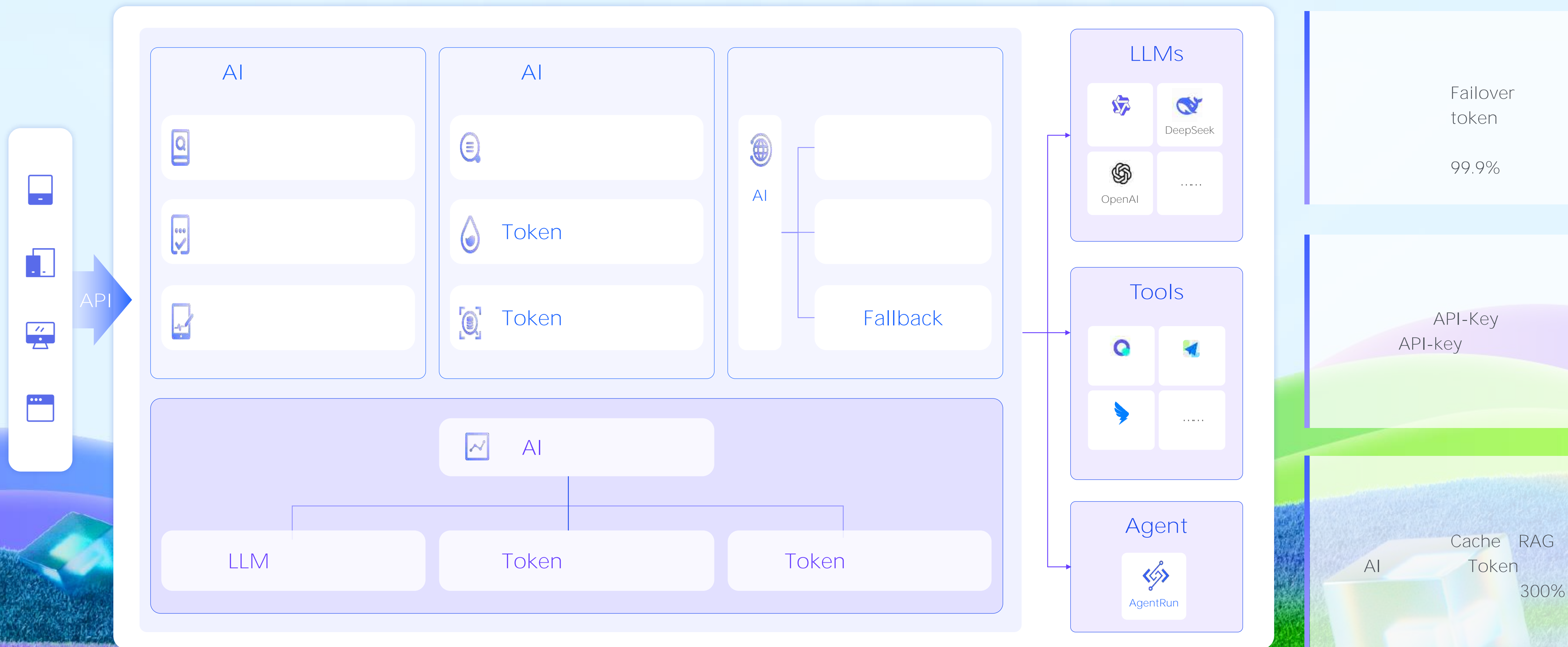
减少 GPU 碎片, 优化稀缺资源利用率

# AgentRun

## AgentRun Gateway

AI

Rest to MCP



# AgentRun

## AgentRun Observability

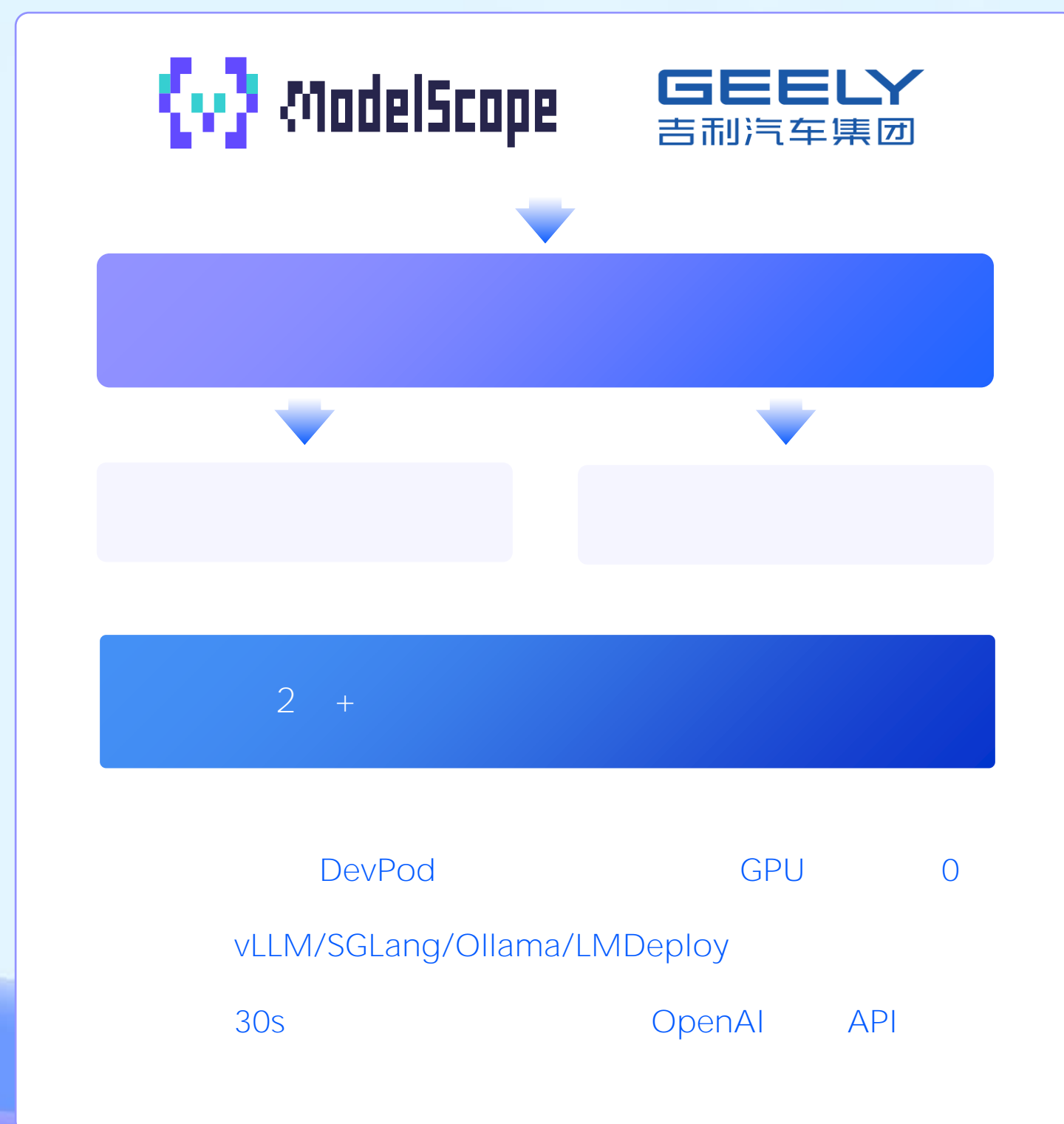
AI



# AgentRun

## FunctionCompute AgentRun Practices

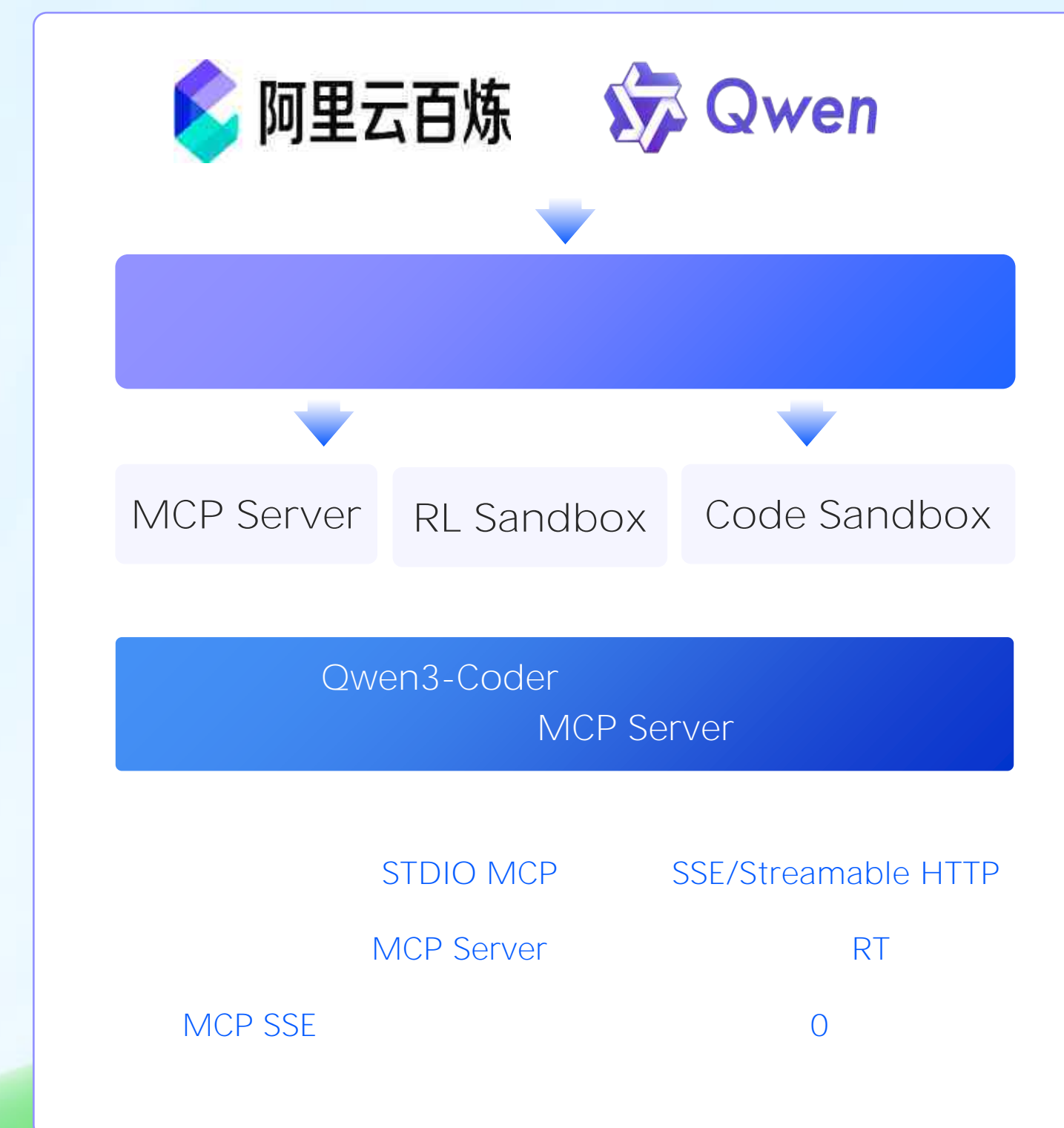
AgentRun



Qwen



AI



AgentRun



# Higress x HiMarket

MCP/Agent

( )

Higress

Maintainer

# CONTENT

01 Higress AI  
AI

---

02 Higress x HiMarket  
Higress x HiMarket

---

03 Higress x HiMarket  
Higress x HiMarket AI

---

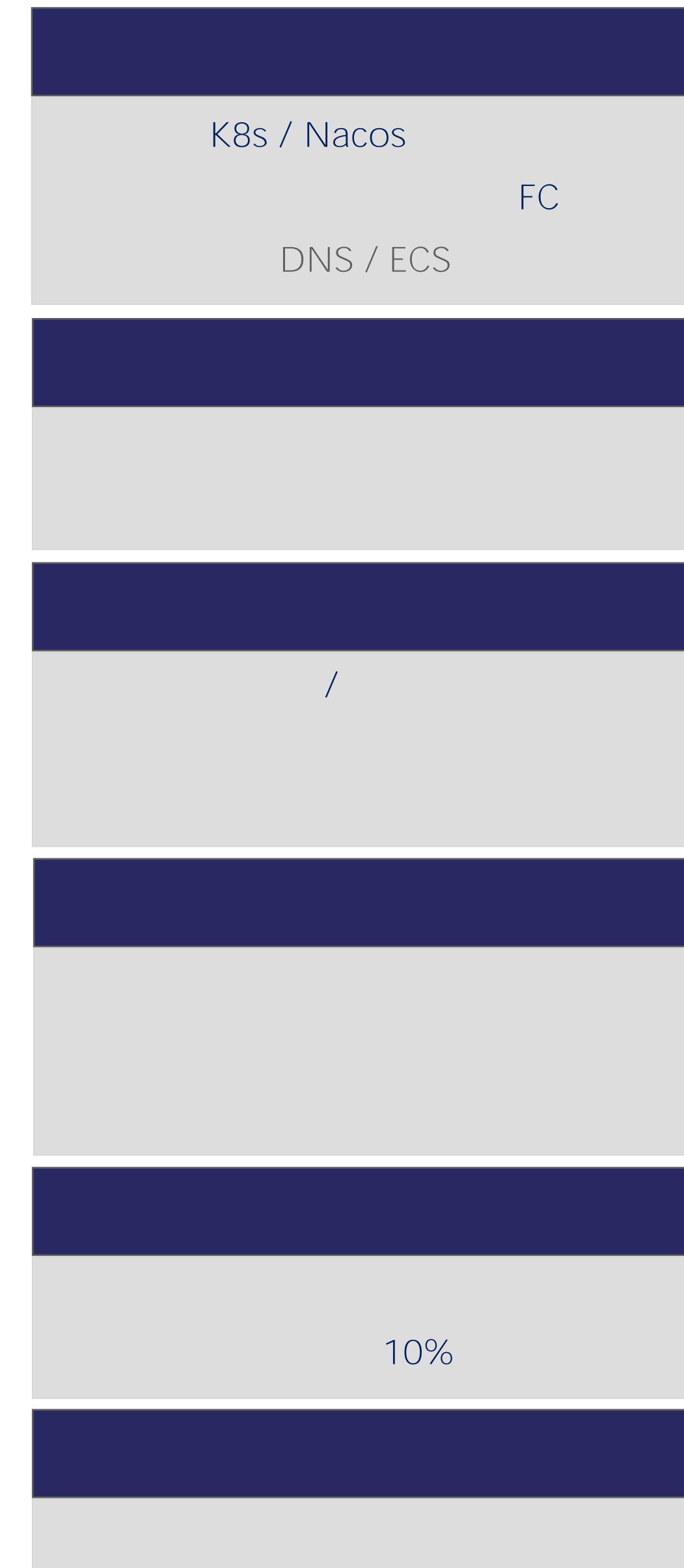
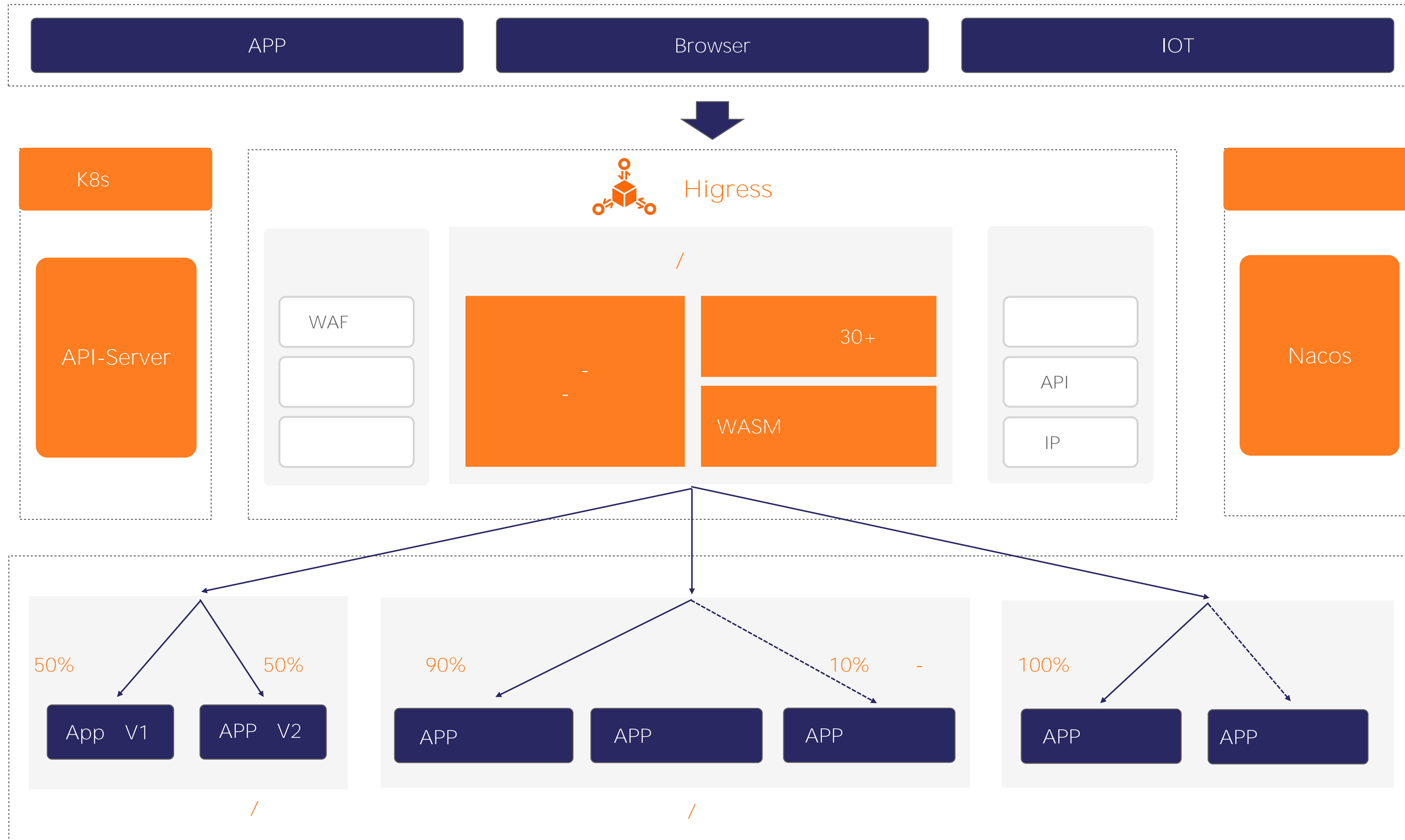
04 Higress x HiMarket  
Higress x HiMarket roadmap

Part 1

Higress AI

AI

# Higress



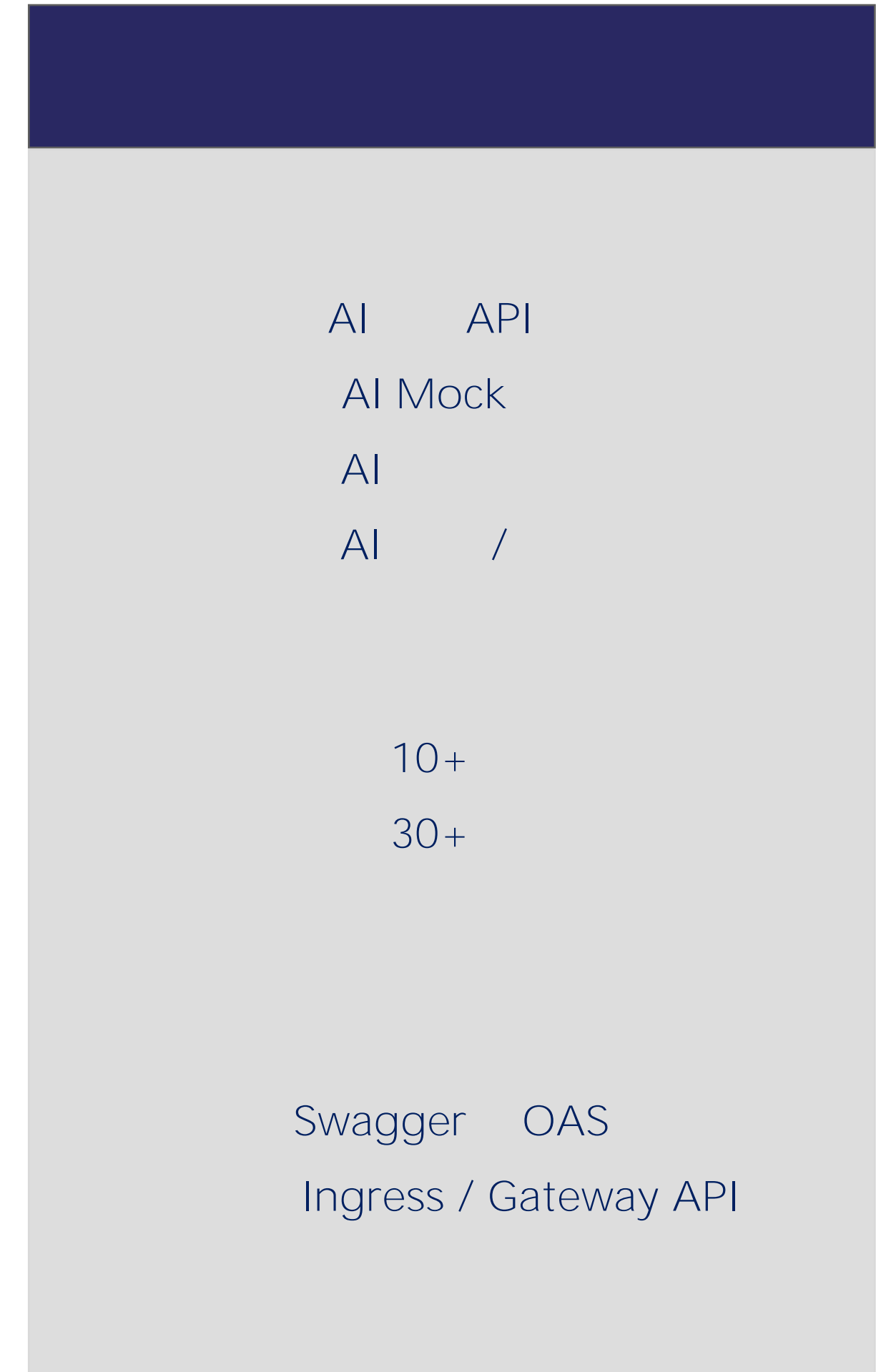
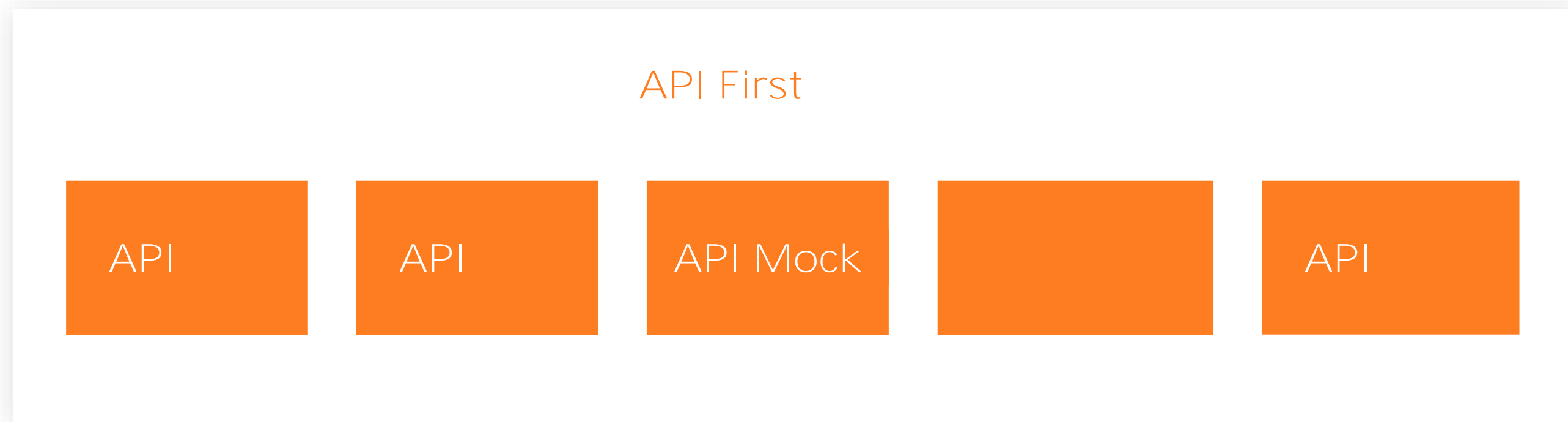
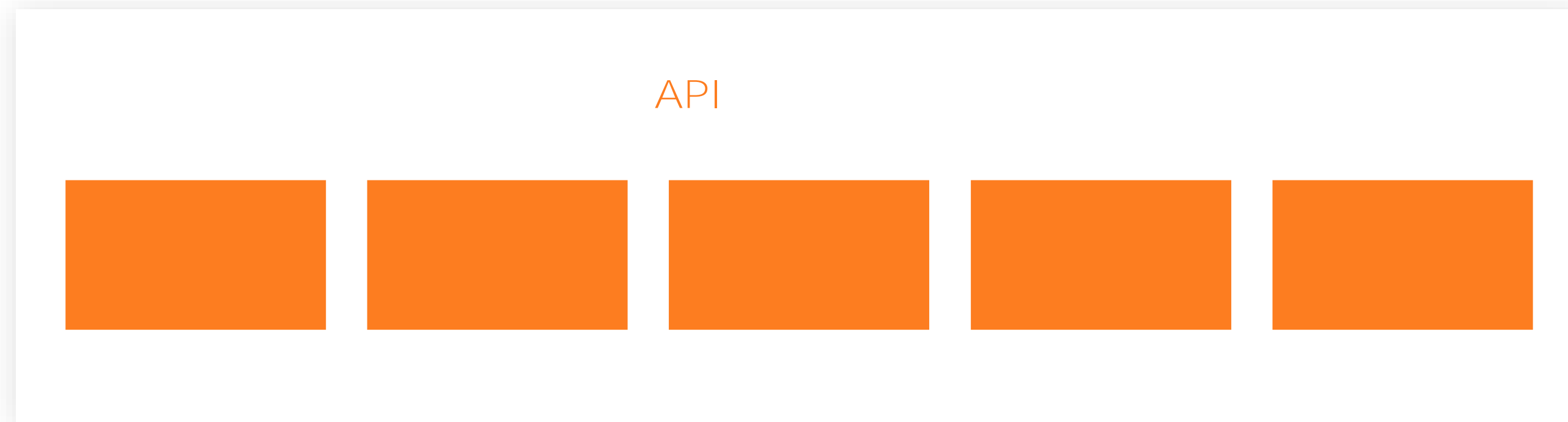
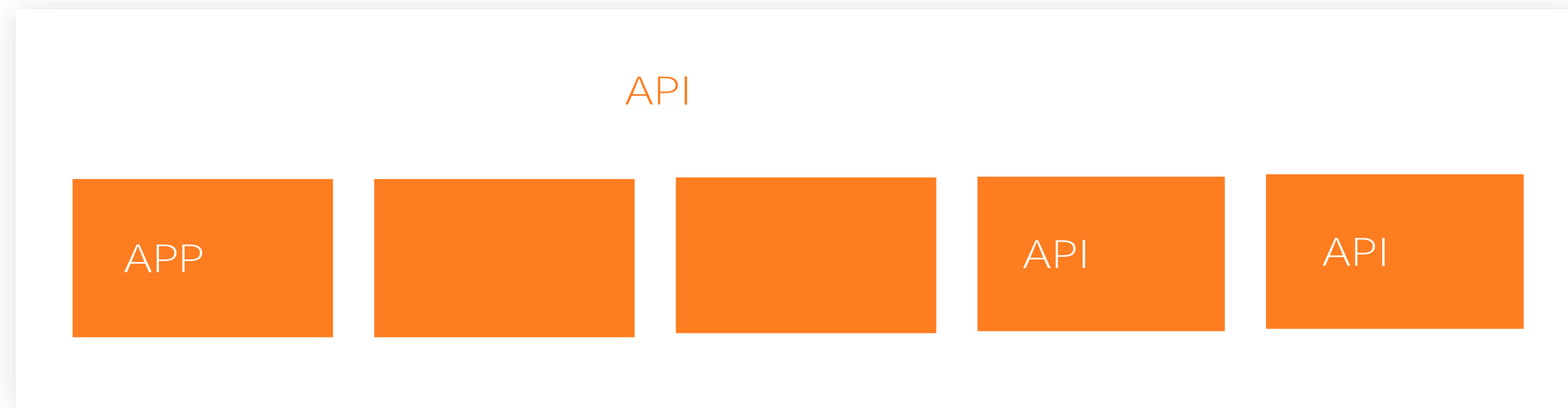
# Higress API

API First

/ API

/

/ API

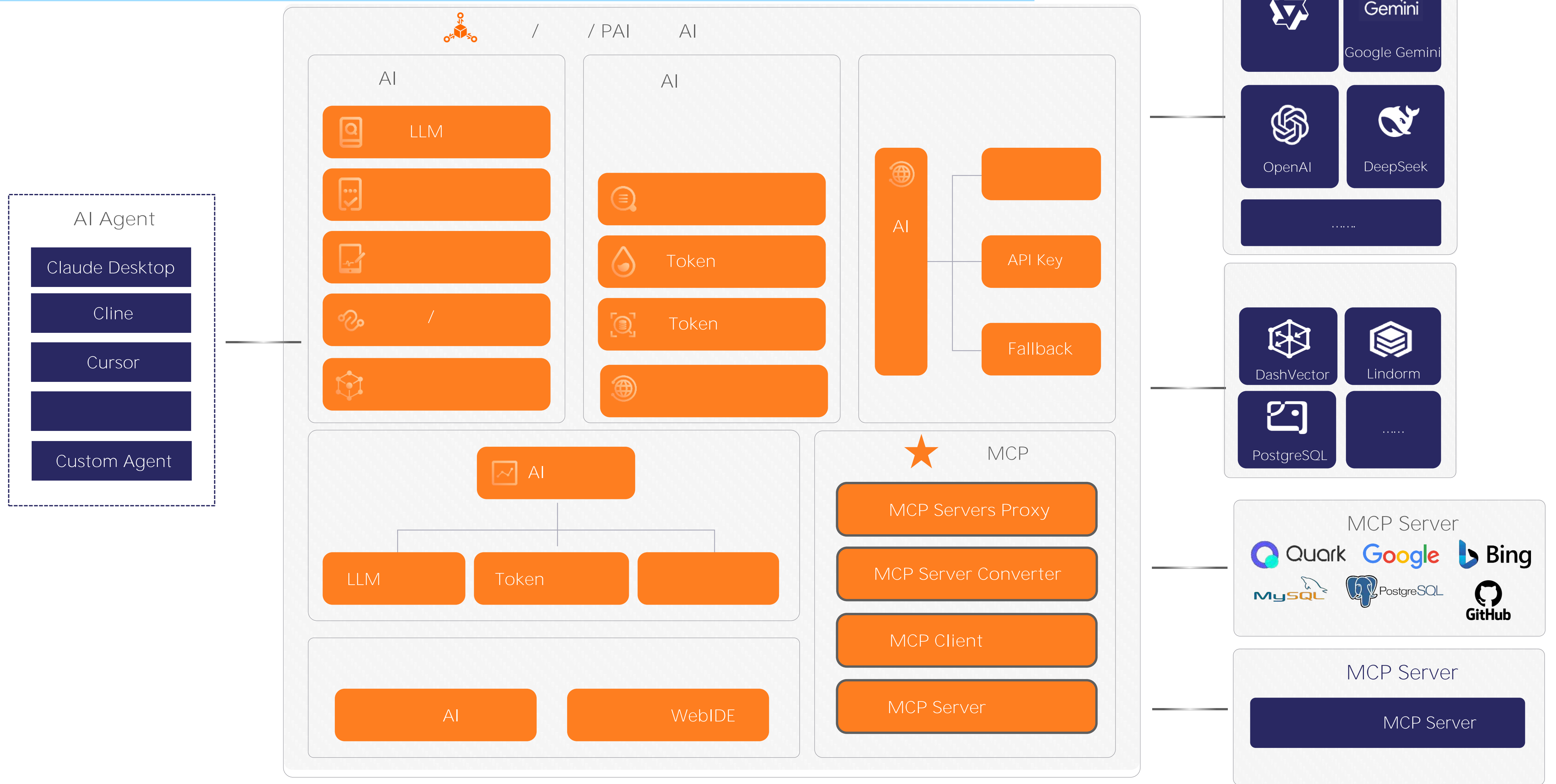


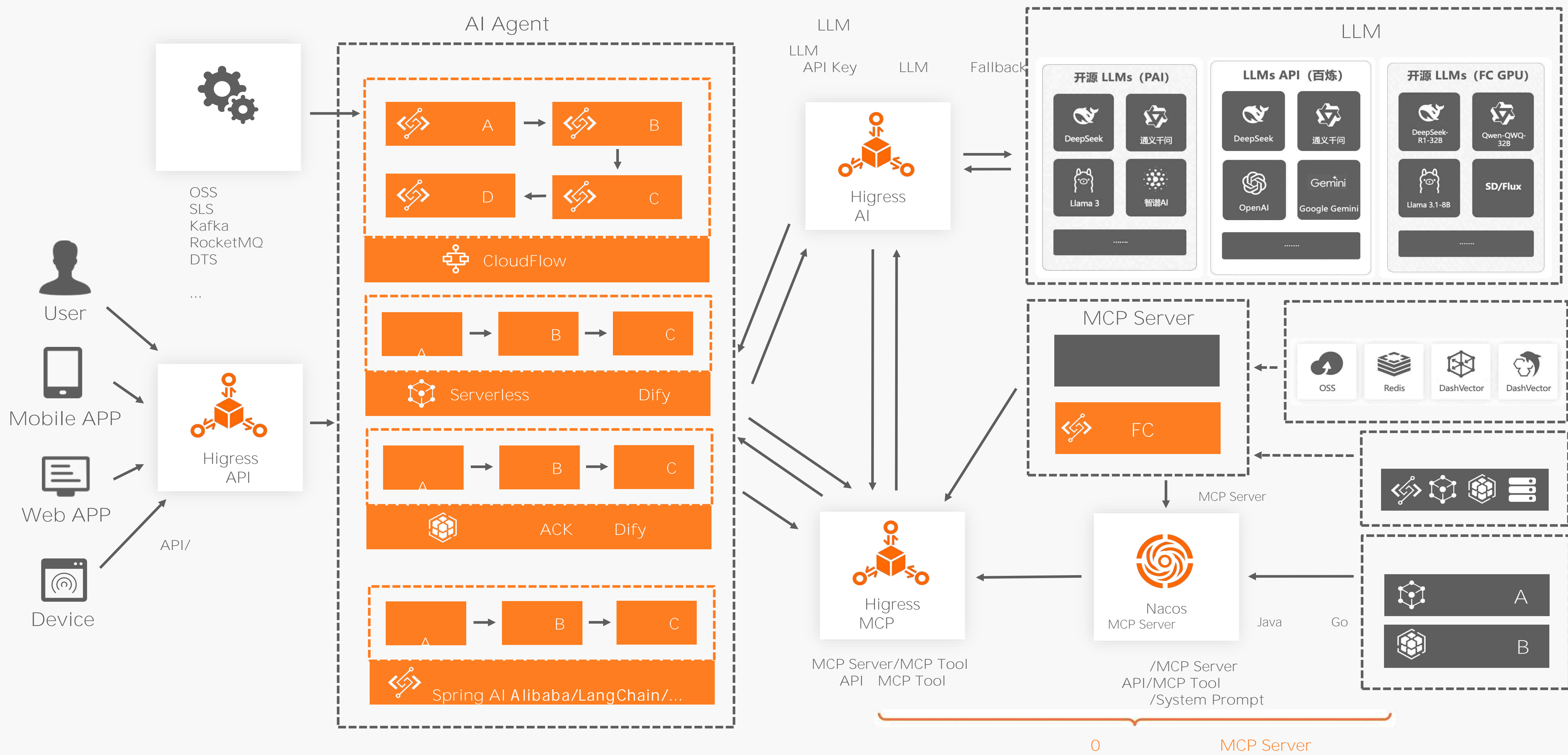
# Higress AI



# Higress AI

## - MCP



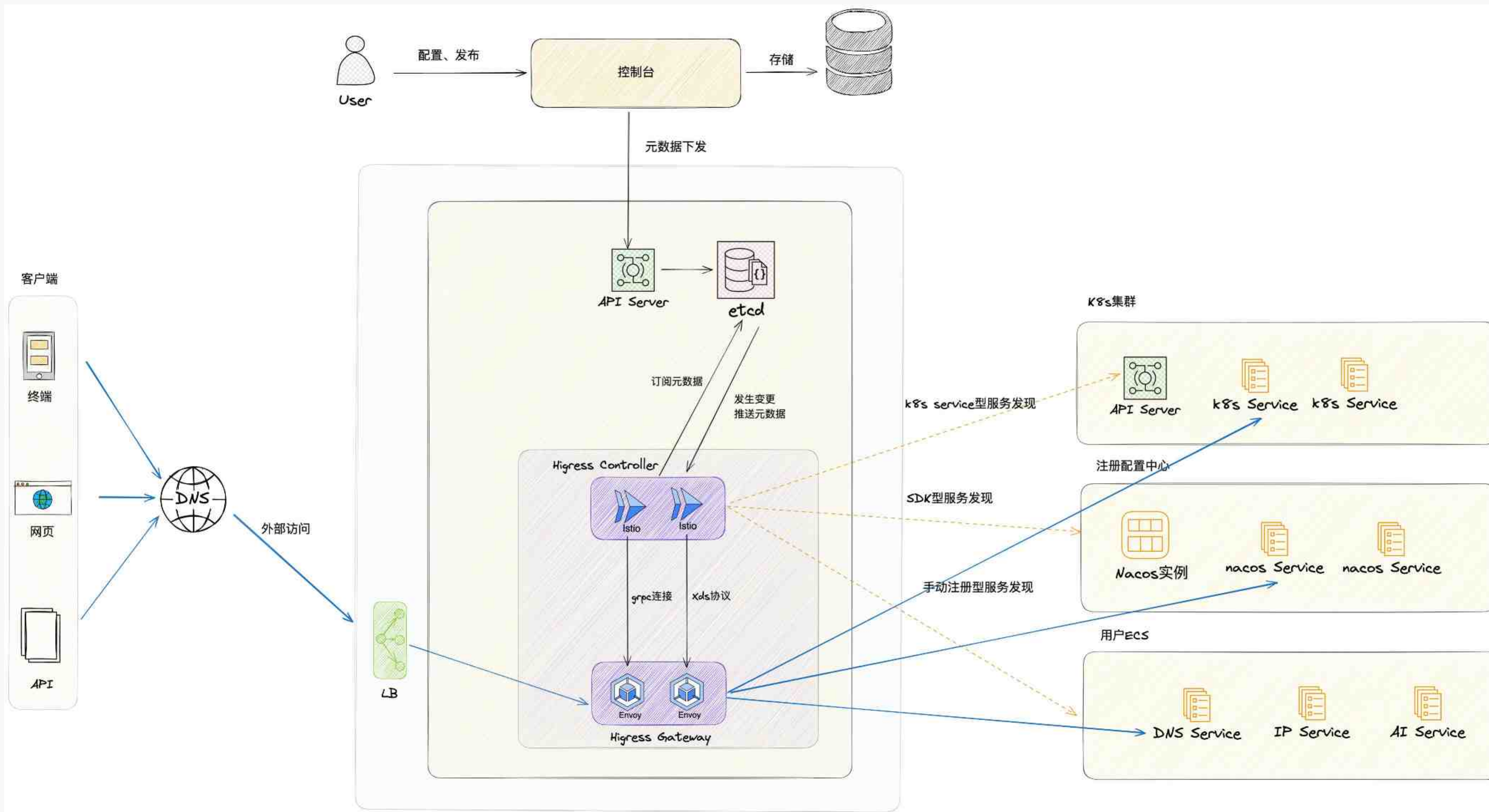


## Part 2

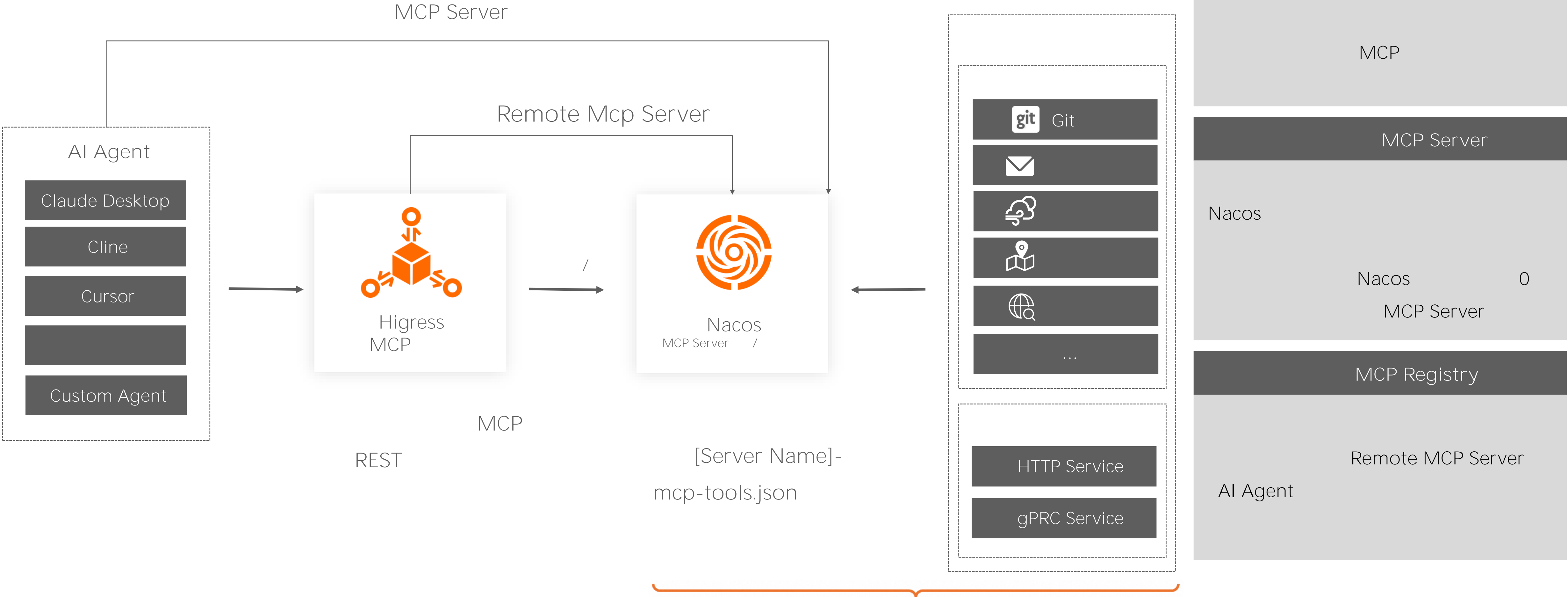
# Higress x HiMarket

Higress x HiMarket

# Higress



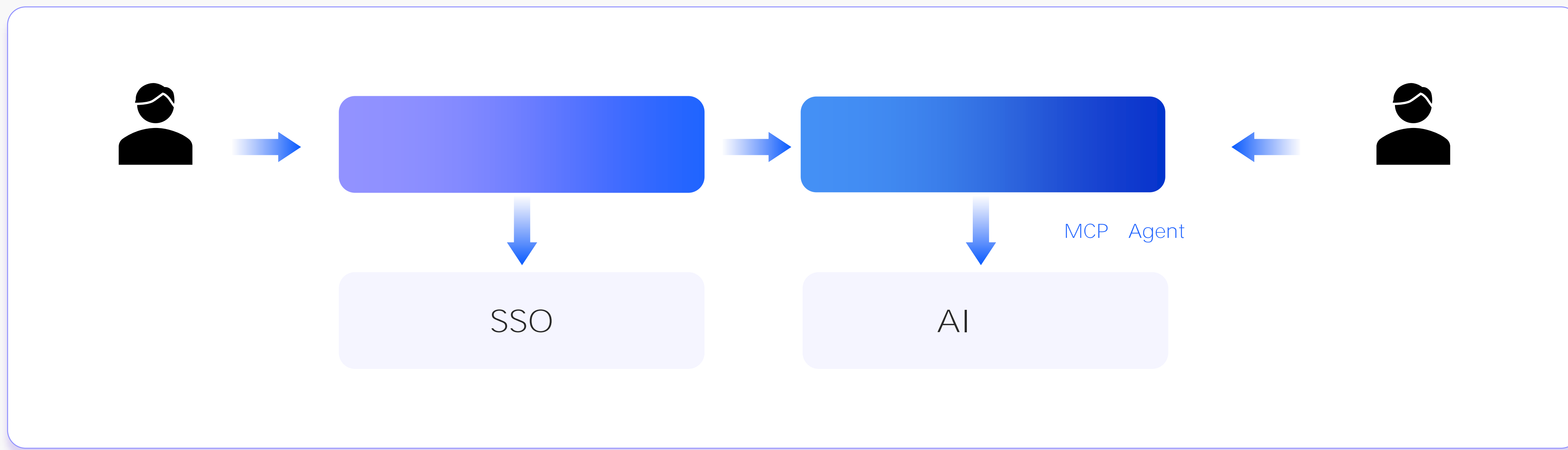
# Higress x Nacos3 MCP



## HiMarket AI

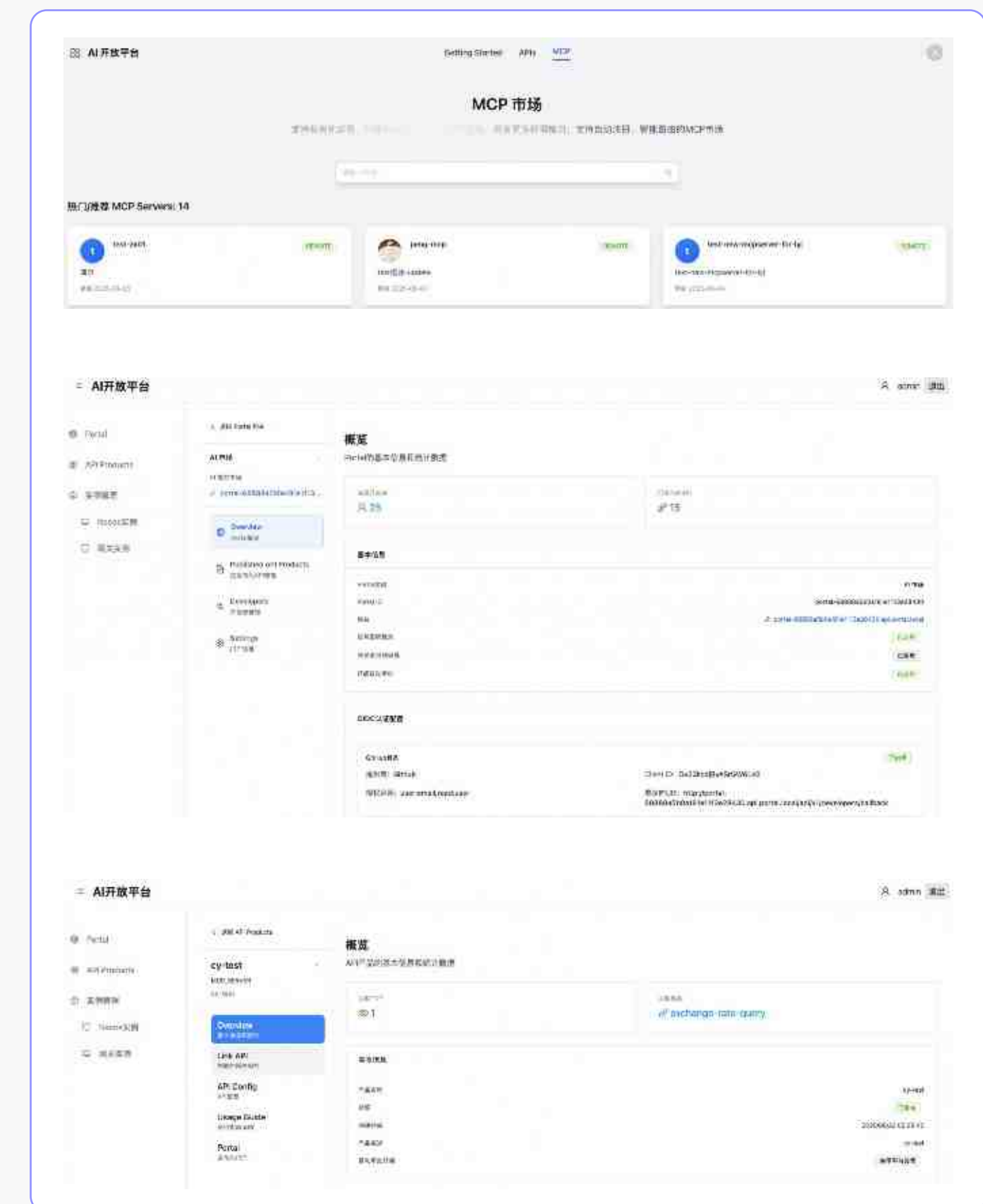
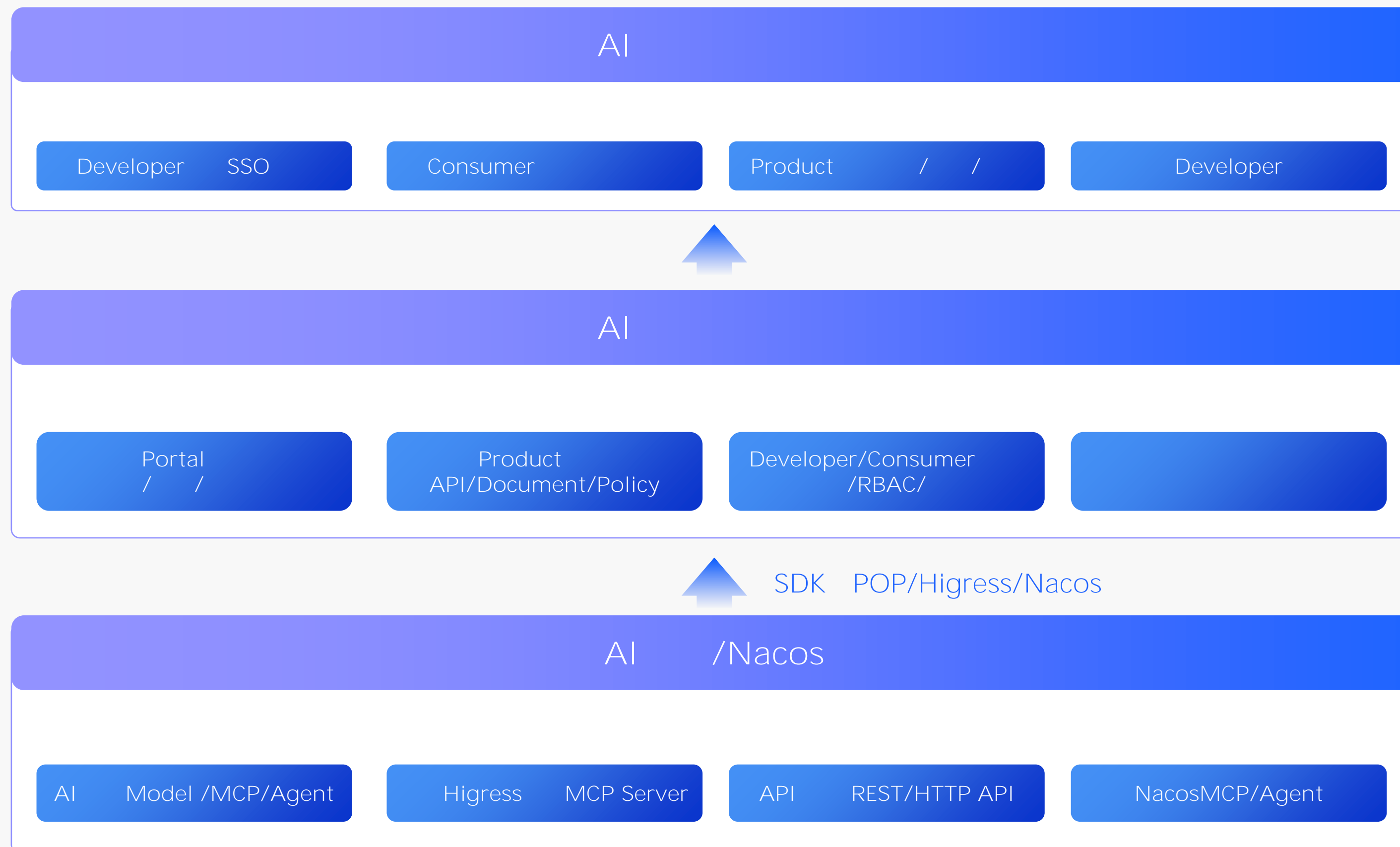
Open Source HiMarket AI Open Platform

AI



## HiMarket AI

Open Source HiMarket AI Open Platform Overall Architecture



## Part 3

# Higress x HiMarket

Higress x HiMarket

AI

# MCP Server

1

2 AI MCP

3 GDPR/CCPA

Developer Tools (8021)

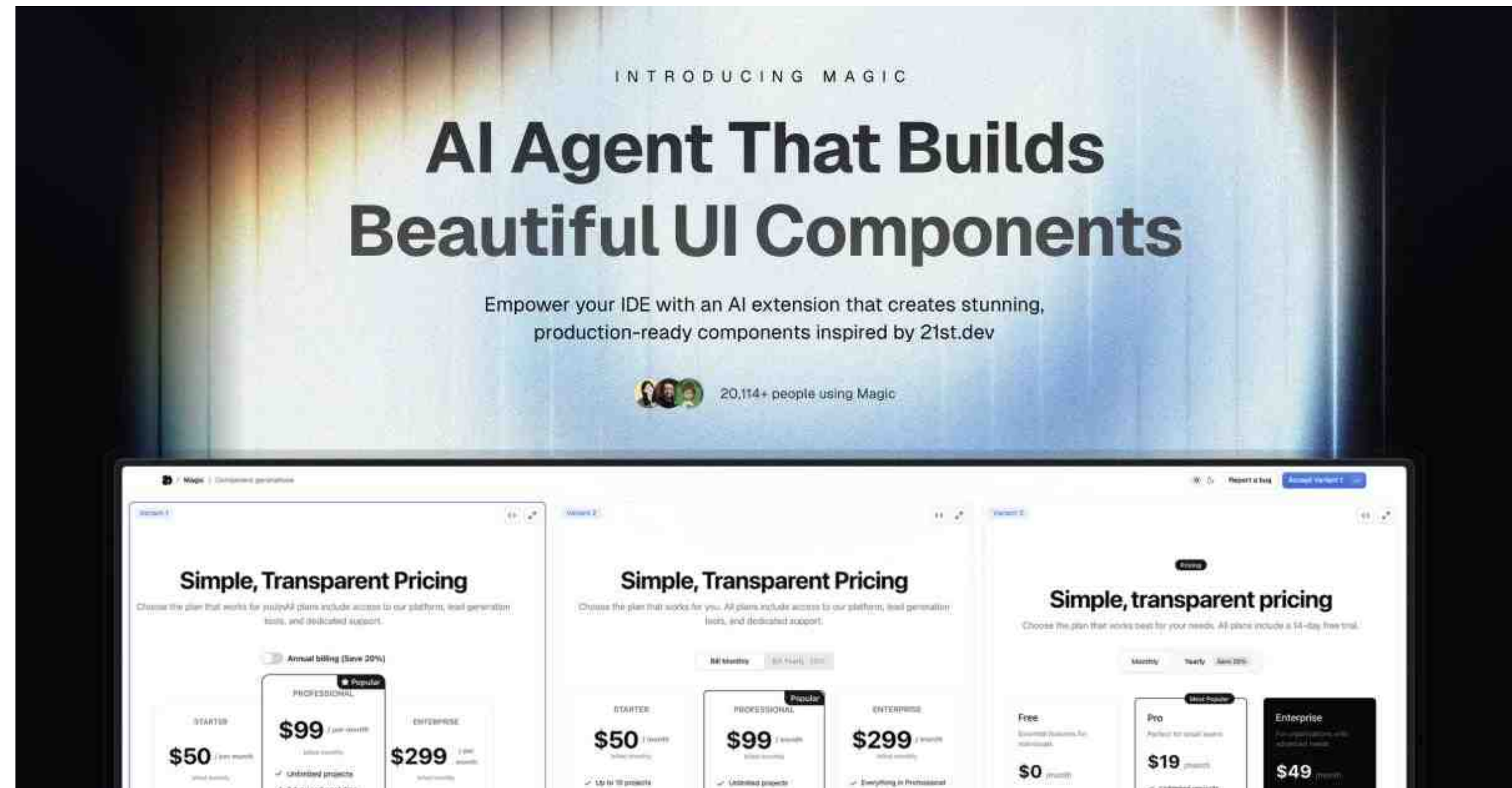
<b>code-mcp</b> Reusable MCP server setup to augment AI coding agents with specific repo guidelines	<b>Memento MCP OAuth...</b> Memento MCP Server with OAuth authentication for Claude.ai	<b>Spring AI MCP Intacct...</b>	<b>FastMCP v2</b> FastMCP server containerized for deployment in Google Kubernetes Engine alongside enhanced-mcp-...
<b>mcp-todo-server</b>	<b>mcp-boilerplate</b>	<b>MCP example Netlify...</b>	<b>MeshLab-mcp</b> mcp server implementation for meshlab
<b>Microsoft Copilot Studio...</b> Silent-office MCP server (silentoffices.com)	<b>Markdown-MCP-NPX</b> NPX wrapper for Microsoft's Markdown MCP server - run without Docker	<b>Mermaid MCP Server</b> Python MCP Server abstracting the official mermaid-cli for ease of use	<b>Jira CLI MCP Server</b> MCP server that wraps jira-cli to enable AI assistants to interact with Jira
<b>MyAnimeList MCP Server</b> MyAnimeList MCP Server	<b>RealtimeRegister MCP...</b> An MCP server for interacting with the RealtimeRegister API	<b>Remote MCP Server on...</b>	<b>MariaDB / MySQL...</b>
<b>Simple MCP Server...</b> Simple MCP Server Implemented in Vertx	<b>MCP Server</b> Test MCP server and client to control advisors api	<b>My MCP Weather Server</b> A dummy MCP server to test the mcp-chat-cli at	<b>MongoDB MCP Server</b>
<b>GenieACS-MCP</b> MCP Server for GenieACS written in Go	<b>Python (MCP) Filesystem...</b> This is a robust Python-based MCP Filesystem Server. It enables AI models and applications to securely...	<b>uMCP</b> uMCP is a server implementation of the MCP protocol, designed for working with the AI in Unity.	<b>Elasticsearch MCP Server</b> Elasticsearch Integration with MCP - Multi-team and multi-environment support
<b>简单 MCP Server 示例...</b> 简单的 MCP Server 示例项目	<b>Hyros MCP Server</b> Hyros MCP Server - Model Context Protocol server providing 24 Hyros API tools for AI agents and automati...	<b>Quickstart Resources</b> Extended the anthropic mcp tutorial to support openai and customGPT connection through http-bridge	<b>Fundamental...</b> MCP server for minecraft

# API

1      API   AI  
Firecrawl      MCP  
AI

2      "      "

3      Agent      API  
MCP Server      Agent      API



**HiMarket** admin 退出

**网关实例**  
管理和配置您的网关实例

Higress 网关 API 网关 AI 网关 专有云 AI 网关

**Higress 网关**  
Higress 云原生网关

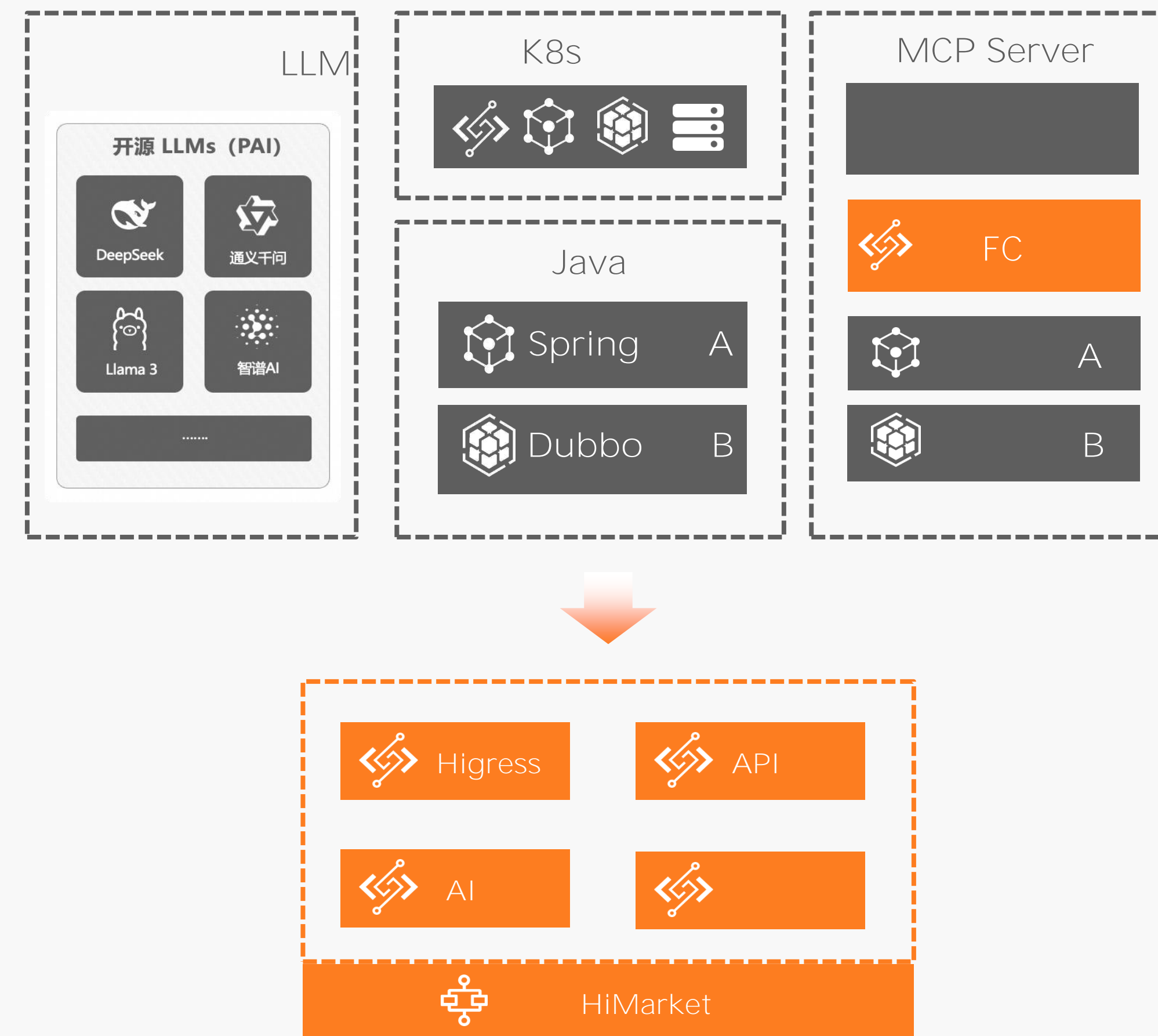
网关名称/ID	服务地址	创建时间	操作
test2 higress-68b5bacfde4b055a680670d54	http://	2025/09/06 03:39:41	删除
test higress-68b5bac7de4b055a680670d53	http://	2025/09/06 03:37:33	删除
demo.higress higress-68b5be03e4b02e0397fb2471	https://	2025/09/02 09:50:59	删除
higress-demo higress-68a08caca218a57e53042ab2		admin 2025/08/16 13:50:36	删除

共 4 条 < 1 > 10 条/页

选择网关类型

- API网关  
阿里云 API 网关服务
- AI网关  
阿里云 AI 网关服务
- Higress网关  
Higress 云原生网关
- 专有云 AI 网关  
专有云 AI 网关服务

取消 确定



HiMarket      Higress      API/AI      AI

< 返回

AI 市场

Overview  
Portal概览

Products  
已发布的API产品

Developers  
开发者管理

Security  
安全设置

Domain  
域名管理

Dashboard  
监控面板

### 概览

Portal概览

#### 基本信息

编辑

Portal名称:	AI 市场	Portal ID:	portal-68888a5b0a181e1113e28430
域名:		账号密码登录:	已启用
开发者自动审批:	已启用	订阅自动审批:	已启用
描述:	AI 能力市场		

注册开发者

32

已发布的API

19

#### 绑定域名

\* 域名

https://example.com

\* 协议

HTTPS

取消

确定

HiMarket

HiMarket

OIDC OAuth2

# API Product



HiMarket admin 退出

< 返回

ChatPPT

- Overview 产品概览
- Link API API关联
- Usage Guide 使用指南
- Portal 发布的门户
- Dashboard 实时监控和统计

### 概览

API产品概览

#### 基本信息

[编辑](#)

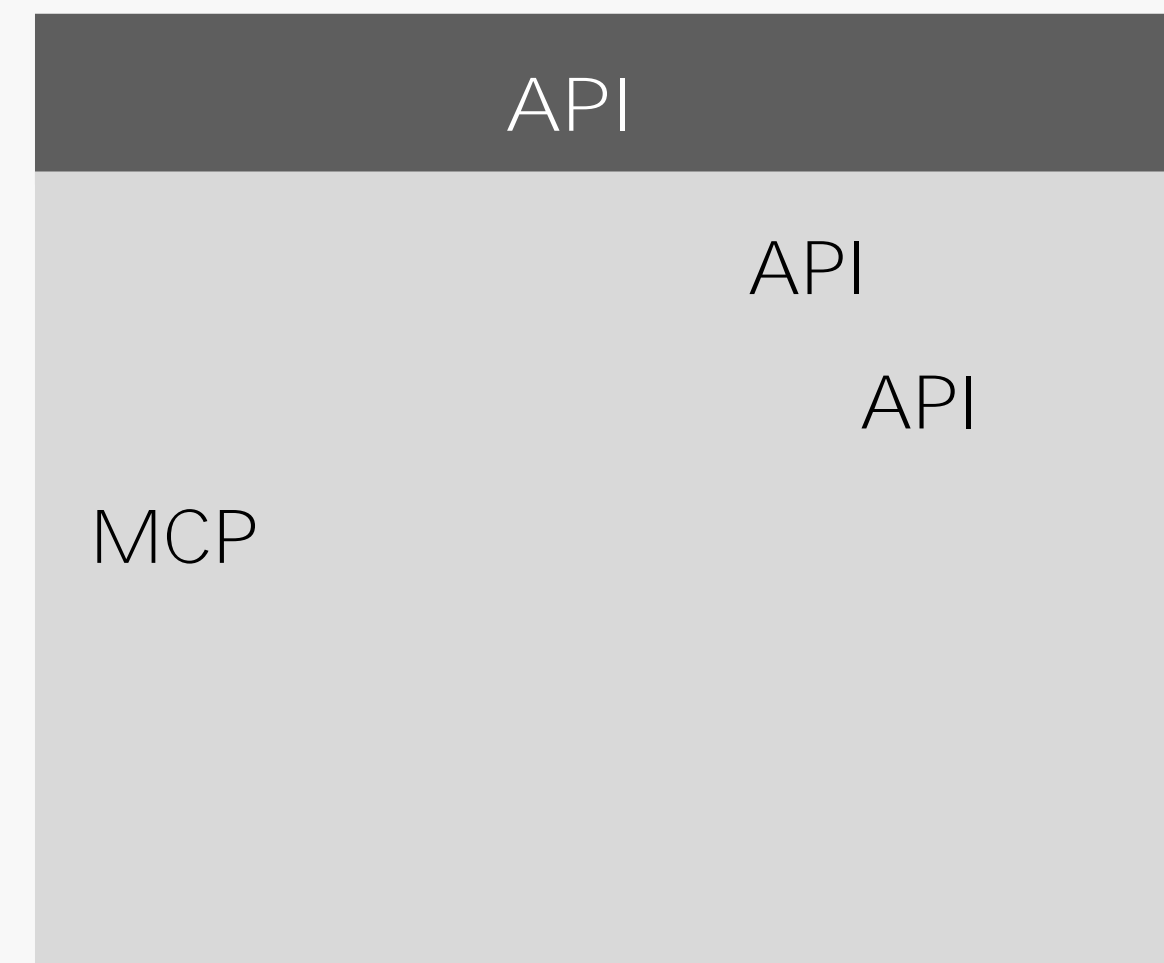
产品名称:	ChatPPT	产品ID:	product-689aa19fe4b0e925c24ebea6
类型:	MCP Server	状态:	已发布
自动审批订阅:	已关闭	创建时间:	2025/08/12 02:06:23
描述:	ChatPPT 提供智能文档的处理能力, 使用 AI 进行 PPT 创作		

发布的门户 0

关联API

exchange-rate-query

订阅用户 0



HiMarket admin 退出

< 返回

ChatPPT

- Overview 产品概览
- Link API API关联
- Usage Guide 使用指南
- Portal 发布的门户
- Dashboard 实时监控和统计

### 发布门户

管理API产品发布的

+ 发布到门户

#### 发布到门户

门户信息	订阅自动审批
<input type="checkbox"/> ai-portal-show portal-68f882b6e61258b9f8251fe9	已开启
<input checked="" type="checkbox"/> AI 市场 portal-68888a5b0a181e1113e28430	已开启

[取消](#) [发布](#)

HiMarket Getting Started APIs MCP 👤

**REST Product**  
2025.08.12 updated

REST API Product

● 未订阅 [管理订阅](#)

Overview OpenAPI Specification

# 天气查询 API 文档

## 接口说明

🔔 快速开始

cURL调用示例

```
curl -X GET \  
  'http://www.test.com/2' \  
  -H 'Accept: application/json' \  
  -H 'Content-Type: application/json'
```

提供实时天

HiMarket Getting Started APIs MCP 👤

**GitHub**  
2025.08.11 updated

MCP Server for the GitHub API, enabling file operations, repository management, search functionality, and more.

Overview Tools (26)

# GitHub MCP Server

**Deprecation Notice:** Development for this project has been moved to GitHub in the <http://github.com/github/github-mcp-server> repo.

MCP Server for the GitHub API, enabling file operations, repository management, search functionality, and more.

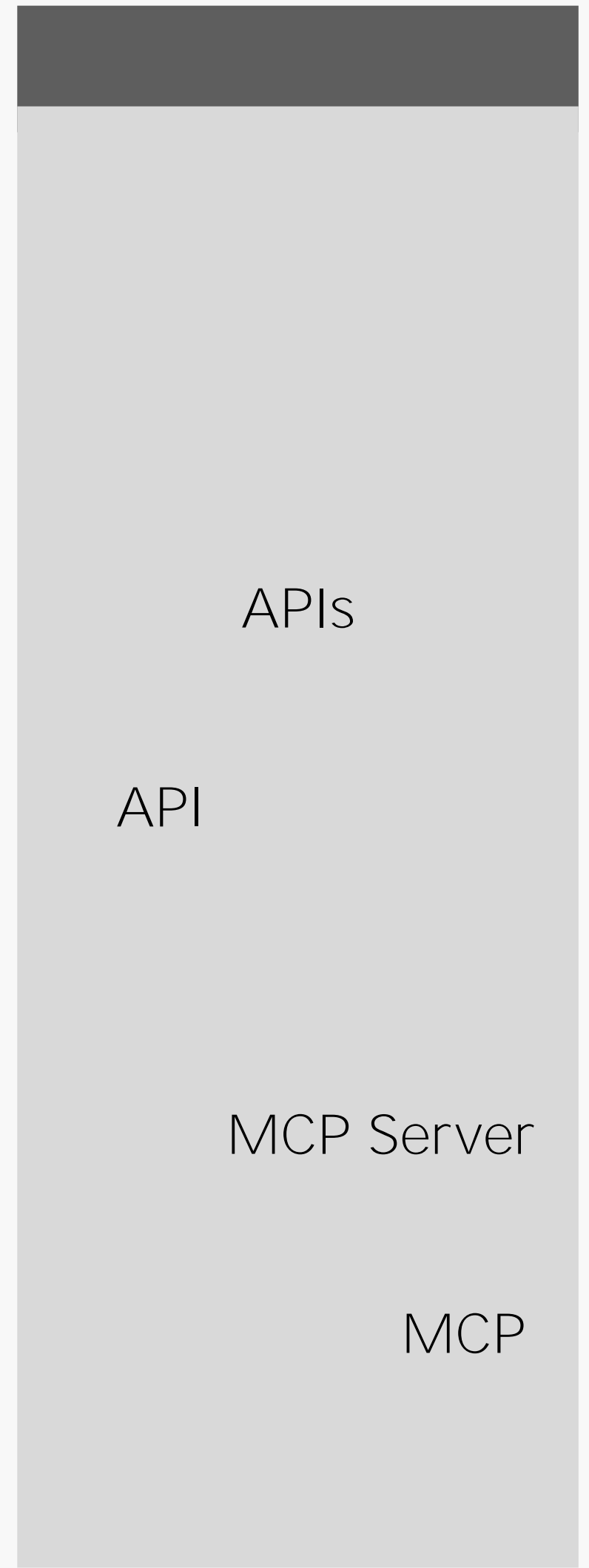
## Features

- **Automatic Branch Creation:** When creating/updating files or pushing changes, branches are automatically created if they don't

### 连接点配置

Stdio

```
{  
  "mcpServers": {  
    "github": {  
      "command": "npx",  
      "args": [  
        "-y",  
        "@modelcontextprotocol/server-github"  
      ]  
    }  
  }  
}
```



HiMarket Getting Started APIs MCP

## MCP 市场

支持私有化部署，兼容MCP市场官方协议，具备更多管理能力，支持自动注册、智能路由的MCP市场

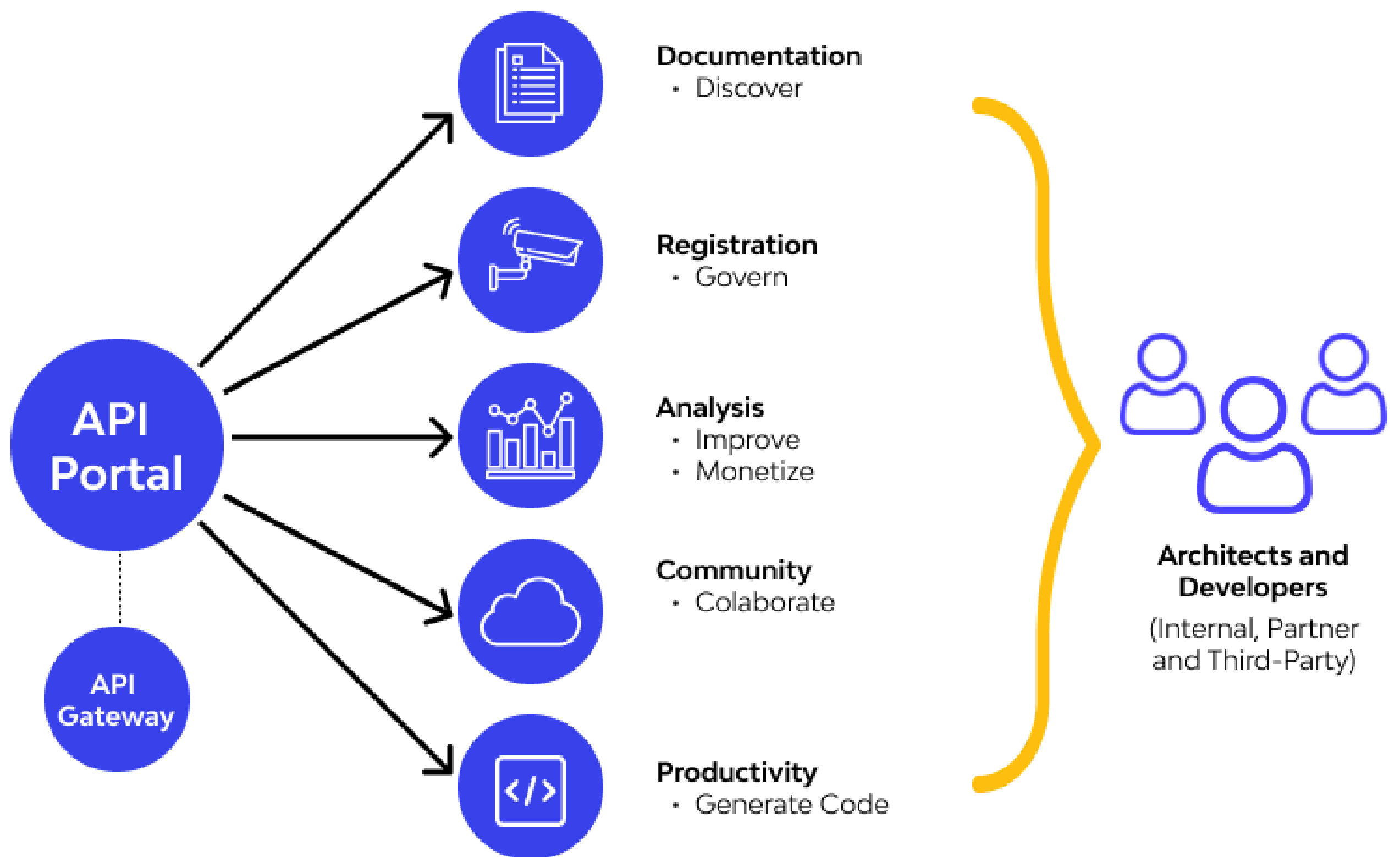
### 热门/推荐 MCP Servers: 15

<b>test-ddd-mcp-liyongjie</b> Remote test-ddd-mcp-liyongjie 更新 2025-10-13	<b>last</b> Remote last 更新 2025-09-24	<b>测试Product</b> Remote 这是一个新的Product 更新 2025-09-21
<b>cy-test</b> Remote cy-test 更新 2025-09-09	<b>Slack</b> Remote Channel management and messaging capabilities 更新 2025-08-12	<b>File system</b> Remote Secure file operations with configurable access controls 更新 2025-08-12
<b>Exchange Rate Query</b> Remote 支持查询中国十大银行的外汇牌价，以及汇率查询 更新 2025-08-12	<b>ChatPPT</b> Remote ChatPPT 提供智能文档的处理能力，使用 AI 进行 PPT 创作 更新 2025-08-12	<b>MSSQL</b> Remote MCP Server for MSSQL database in Python 更新 2025-08-11

## Part 4

# Higress x HiMarket

Higress x HiMarket    roadmap



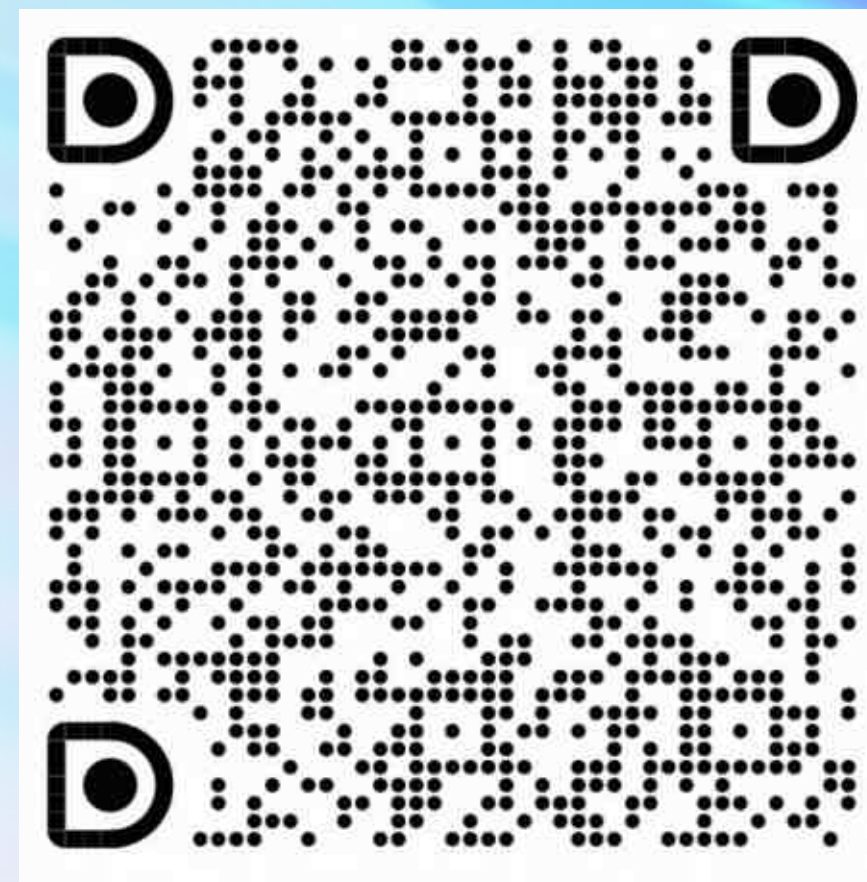
### Agent

1. Nacos Agent
2. Agent
3. Agent DIY

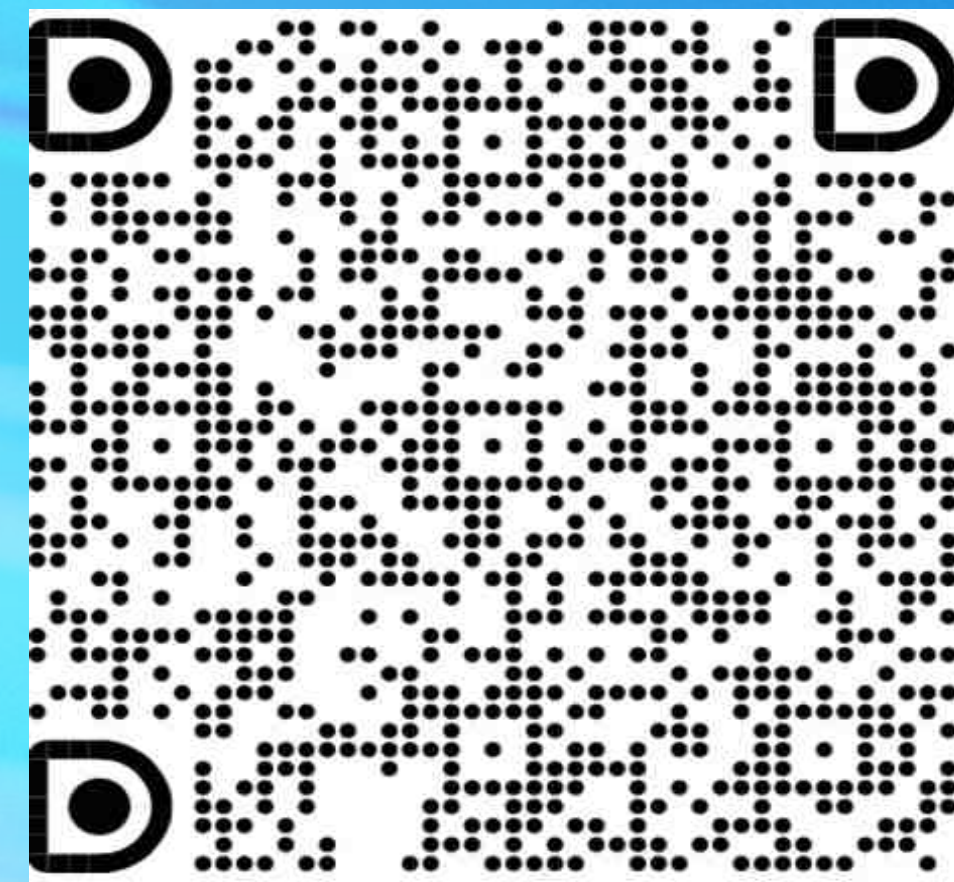
1. API
2. API

1. Portal Consumer Product Developer
- 2.
- 3.

1. API MCP
- 2.
- 3.



HiMarket



Higress

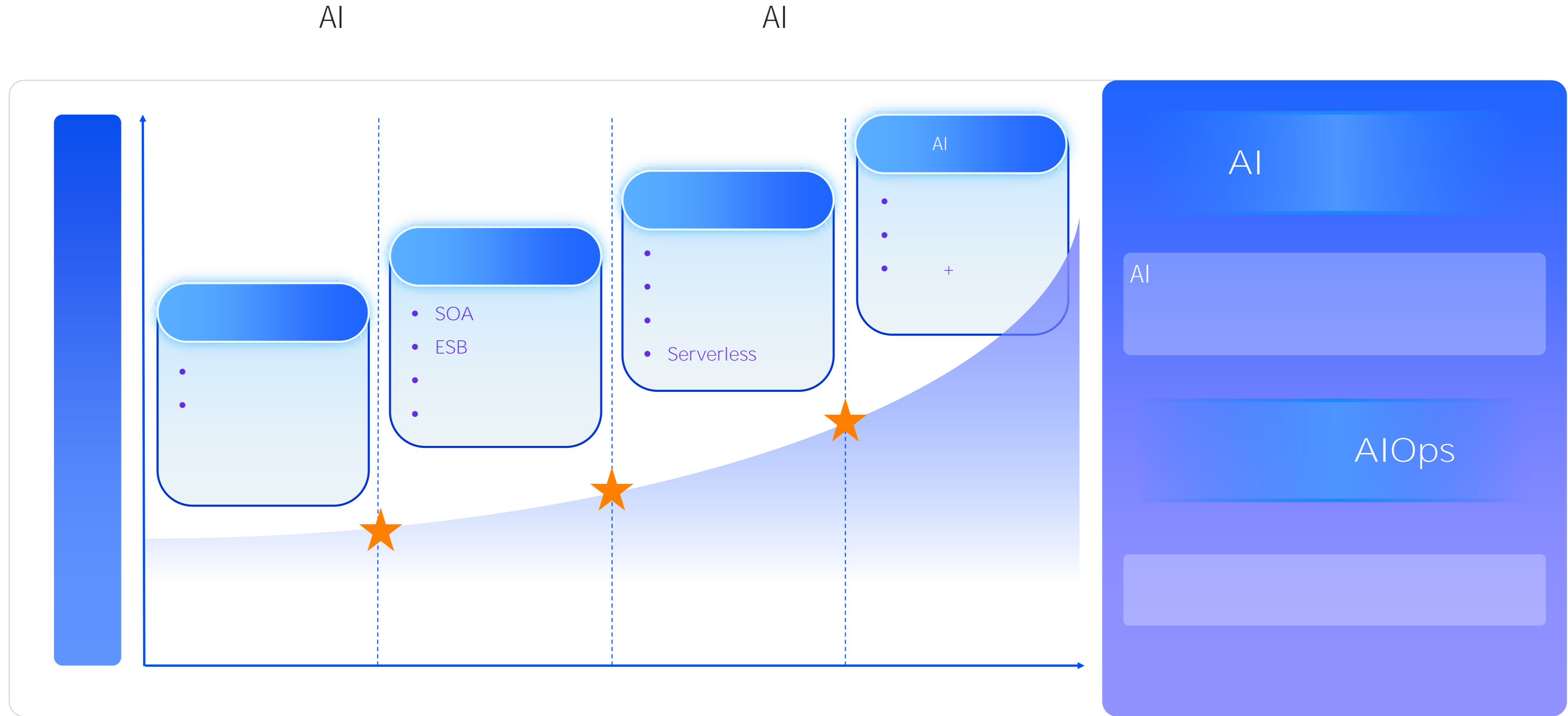
<https://github.com/alibaba/higress>

Rebuilding Observability

AIOps

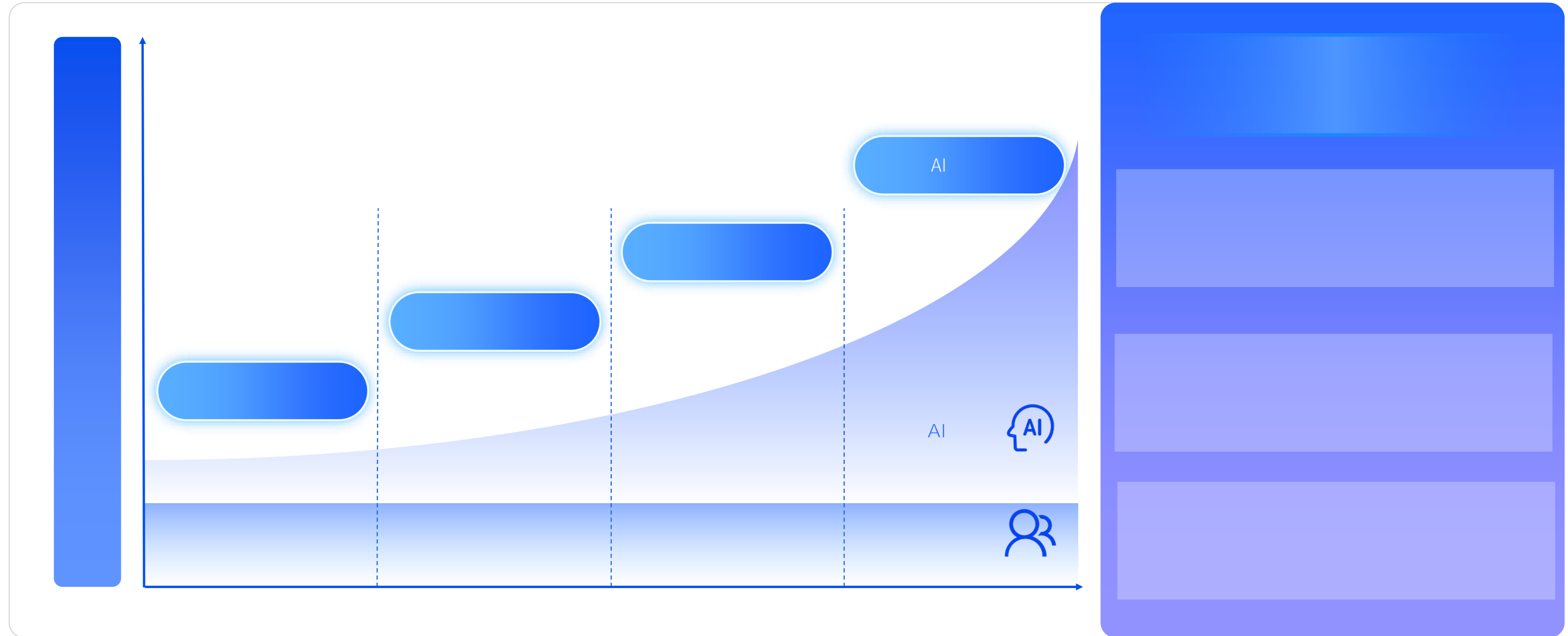
LLM-Powered CloudMonitor 2.0 & New Paradigms for AIOps

### The LLM Era: New Paradigms for Applications & Operations



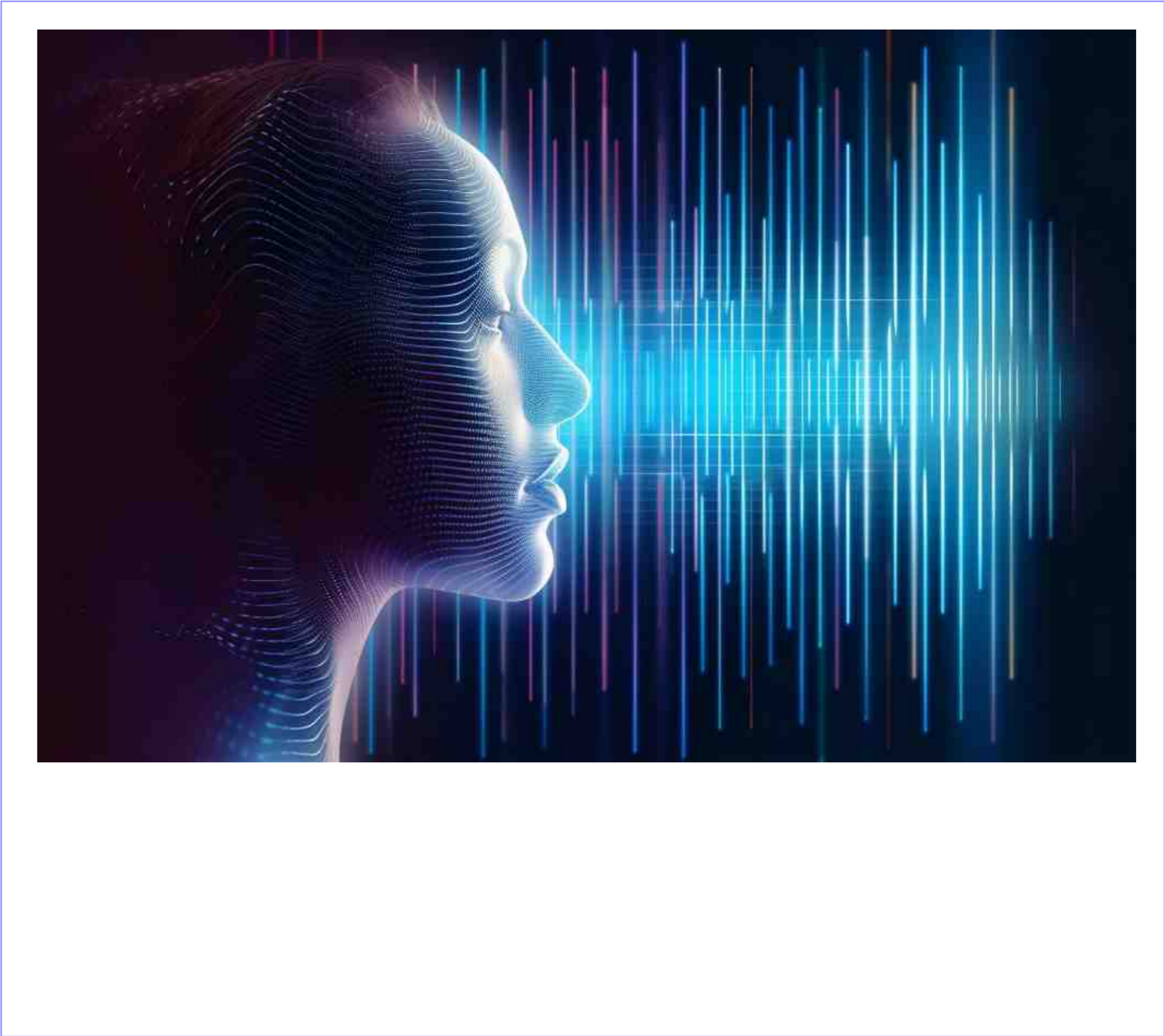
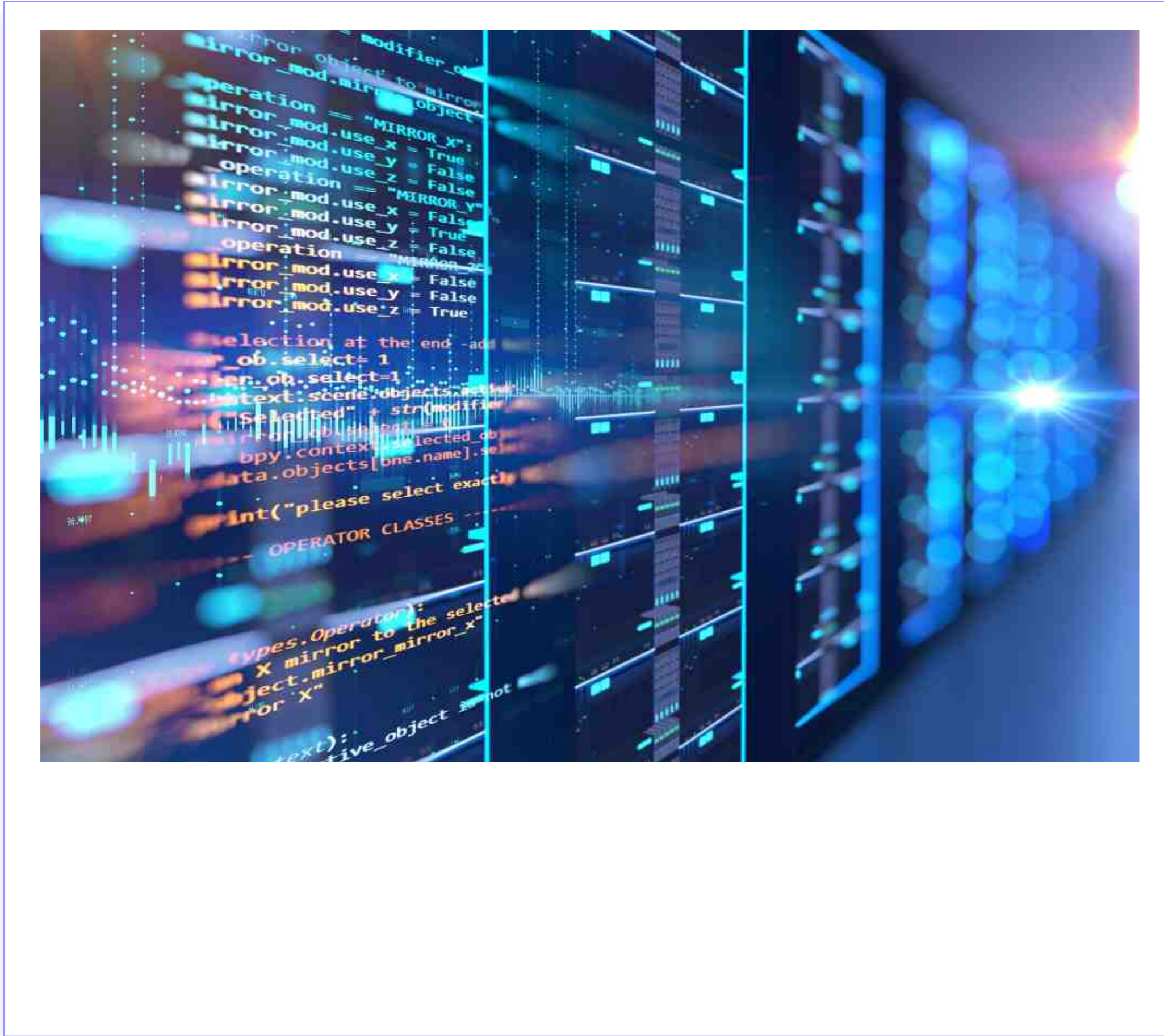
# AIOps

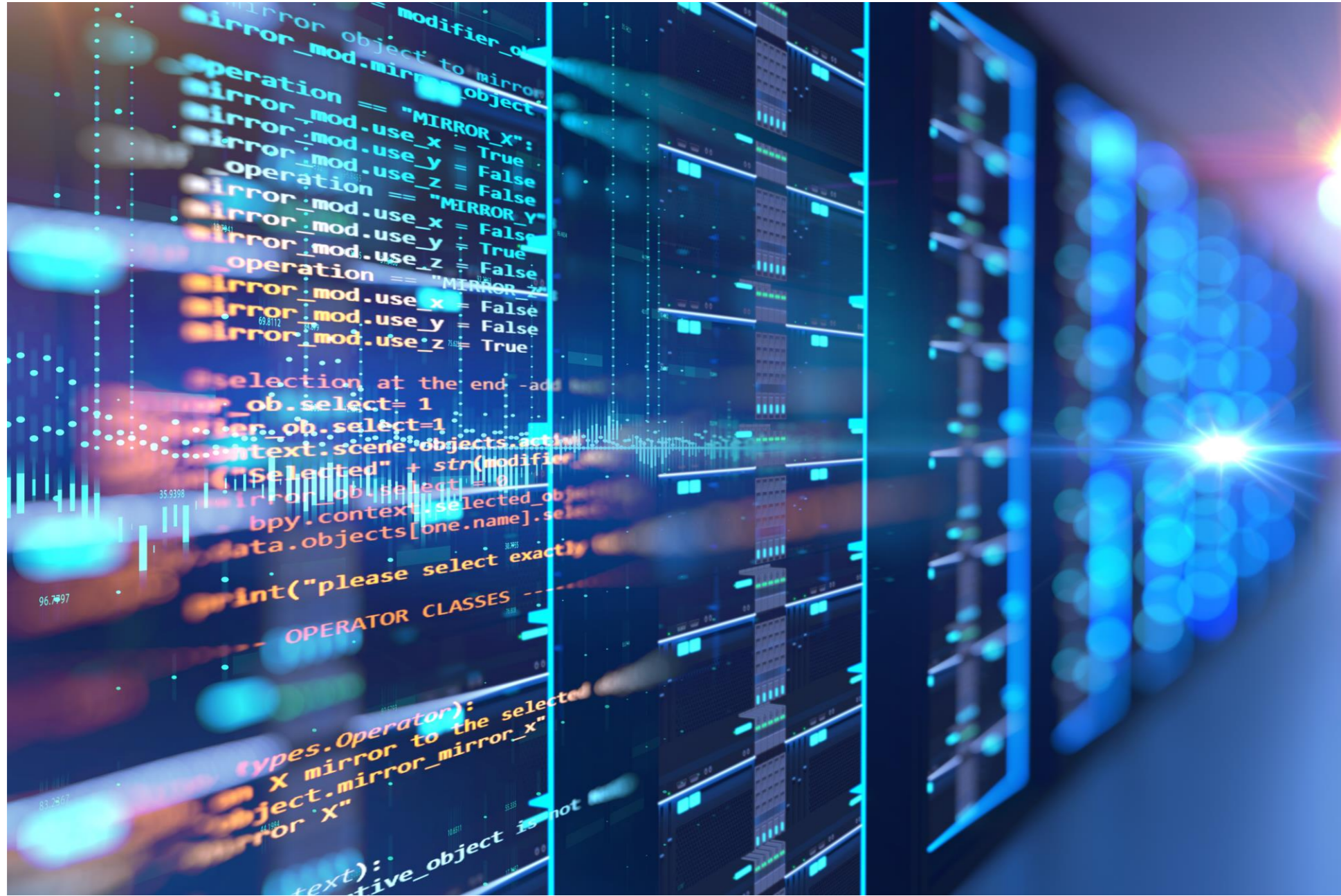
AIOps at the Tipping Point



# AIOps

Two Major Challenges for AIOps in the LLM Era





数据割裂分散，大模型查不到、不好查，如同盲人摸象



数据爆炸增长，平台存不下、存不起，限制大模型有效发挥



让大模型直接处理海量低价值密度的数据，无异于杀鸡用牛刀

Observability Data Platform: Unified Ingestion, Processing, and Storage



每日 EB 级存储规模，综合成本较开源自建下降 50%

×

General Operators × Observability Data Operators: Simplifying Large-Scale Data Analysis





听不懂运维“黑话”，大模型无法精准理解，结果差强人意

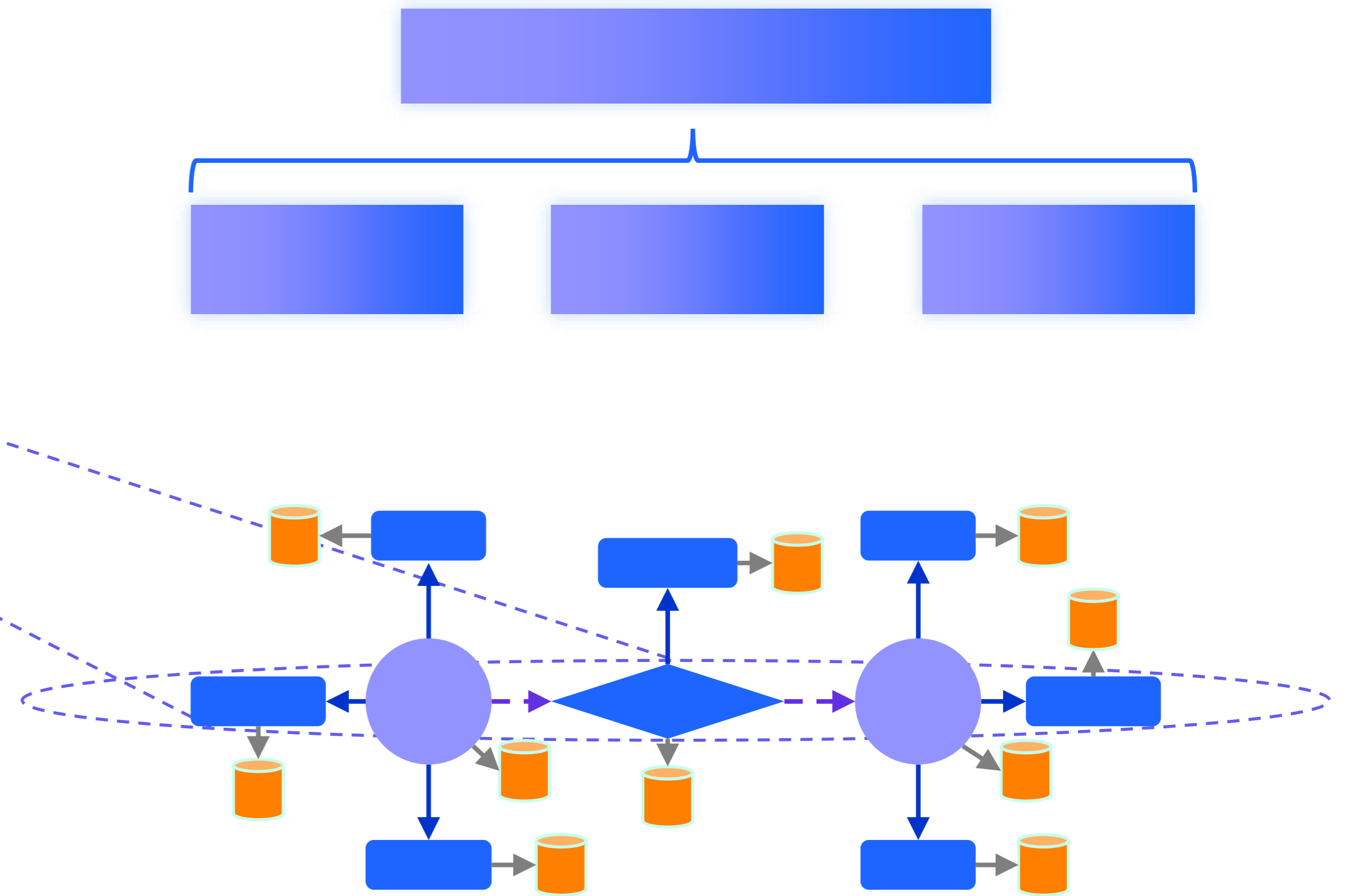
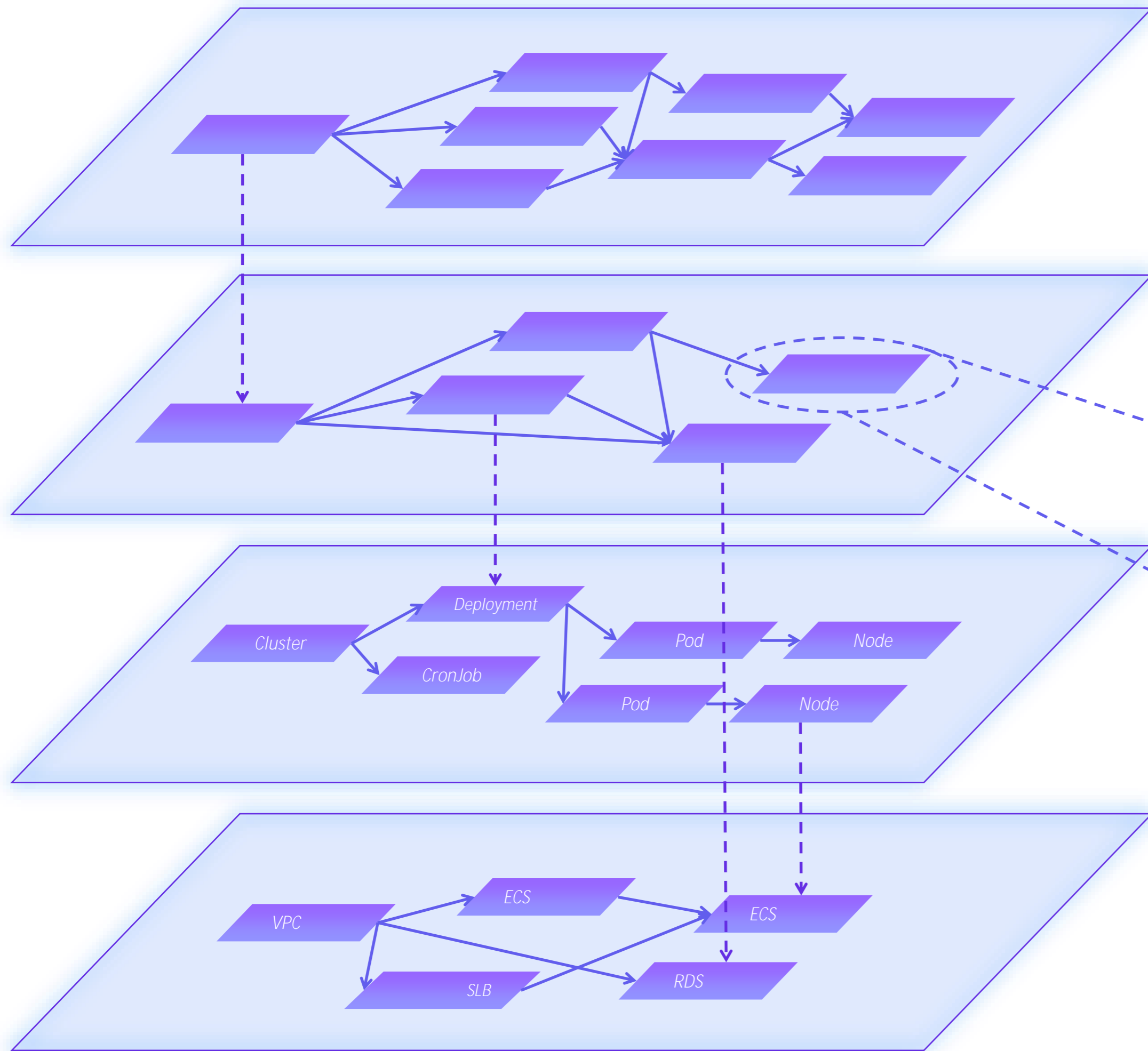
系统间的依赖关系错综复杂，大模型容易只见树木不见森林

数据关系缺失，大模型难以区分逻辑关联，可能推理幻觉

# UModel

Introducing the Unified Model (UModel)

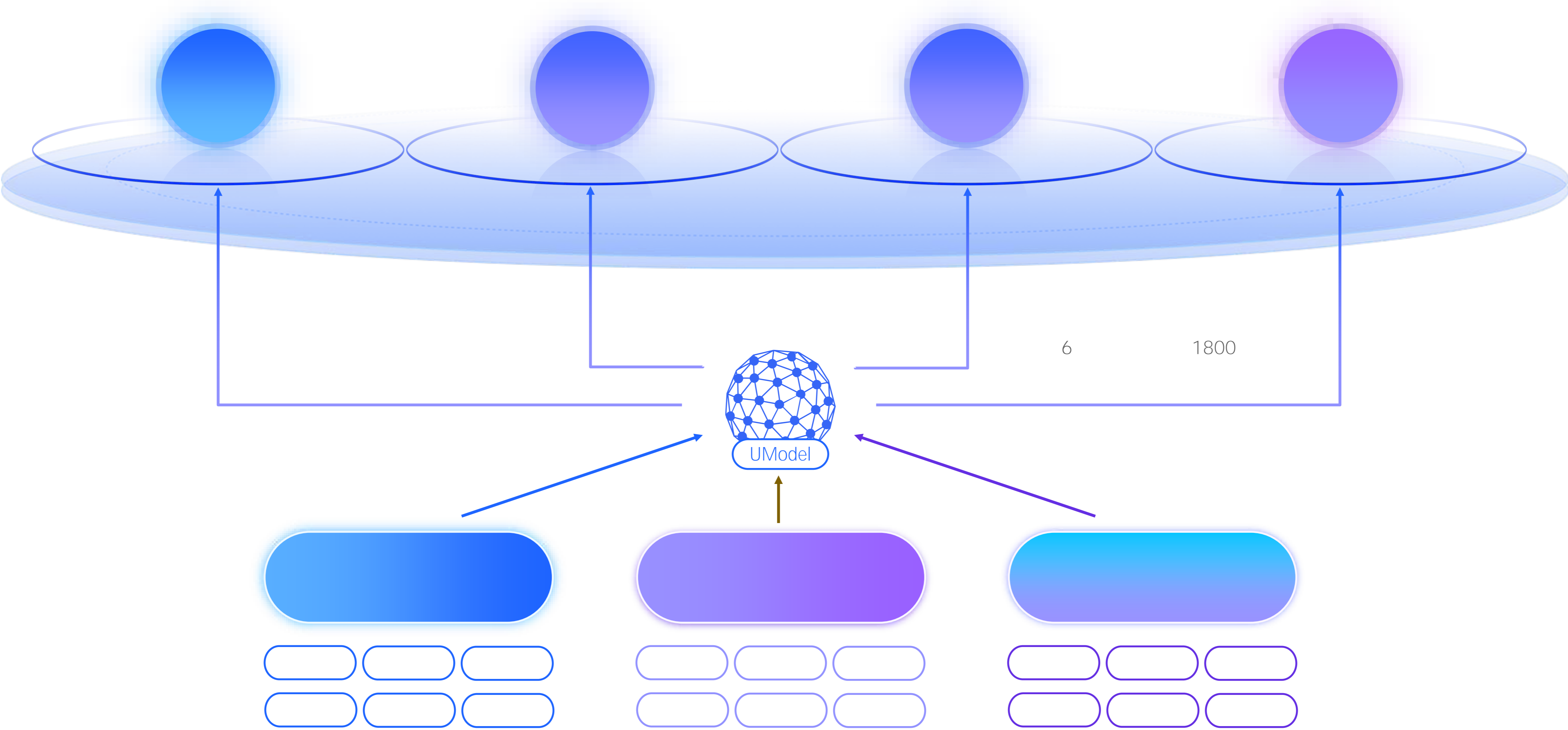
“ ”



UModel

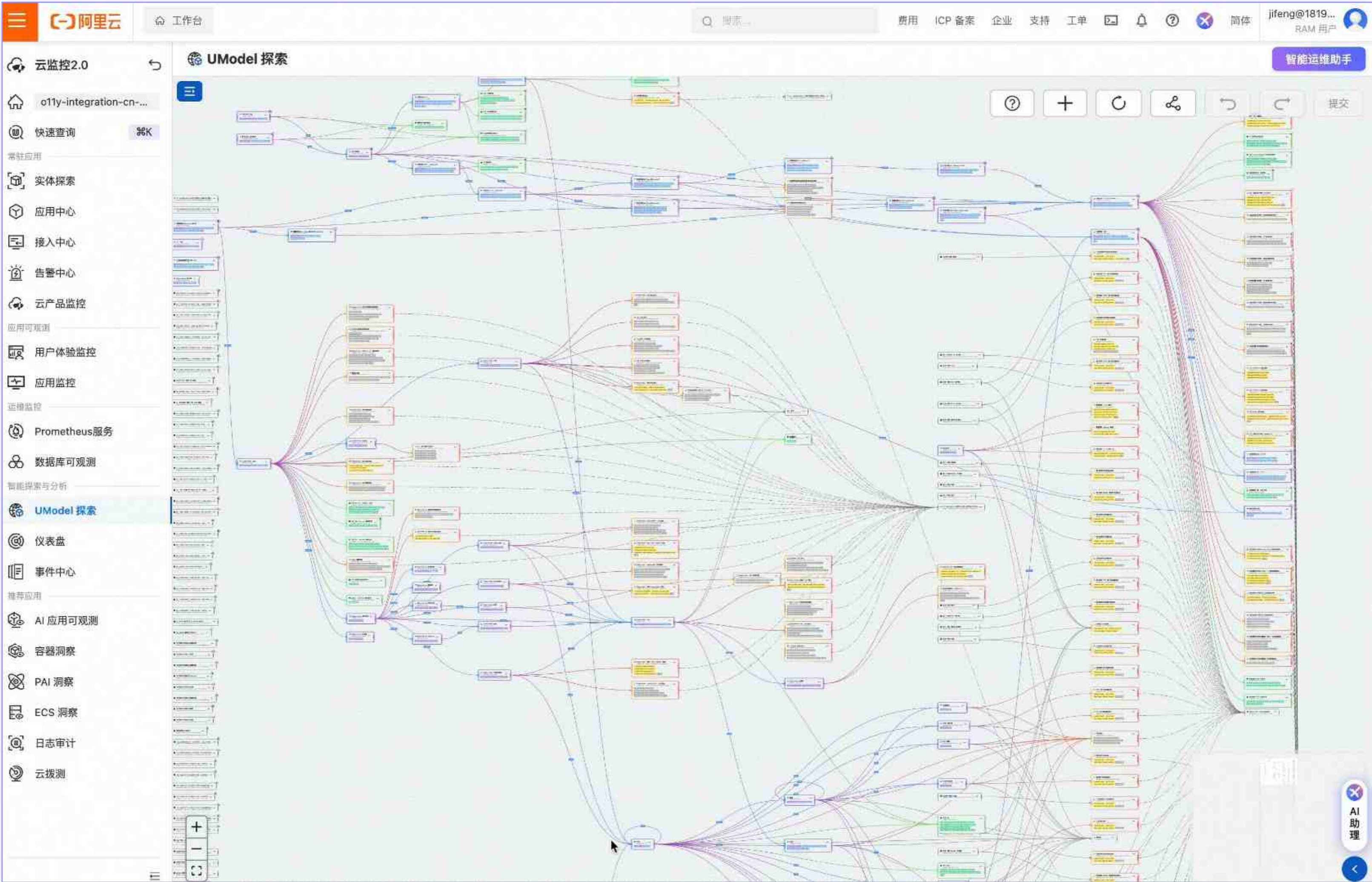
# UModel

Rebuilding Observability Data with the Unified Model (UModel)



# Demo UModel

## Demo: UModel Explorer & Global Entity Topology



# Demo

## Demo: Troubleshooting with the Entity Topology

The screenshot shows the Alibaba Cloud Alert Center console. The main content area displays a table of alerts. The table has columns for severity, trigger time, alert content, associated resource, monitoring type, duration, and actions. A tooltip is shown over one of the alerts, providing detailed information about the alert's content and the specific conditions that triggered it.

严重等级	触发时间	告警内容	关联资源	监控类型	持续时长	规范	操作
严重	2025-09-19 18:23:06	应用提供服务统计 接口名称: /api/order/order/getOrdersByCondi...	service-gateway	应用监控	12分钟 5秒	ser 24	通知历史   详情   查看事件   根因分析
严重	2025-09-19 18:22:06	应用提供服务统计 接口名称: /api/order/order/getOrdersByCondi...	service-gateway	应用监控	13分钟 5秒	ser 24	通知历史   详情   查看事件   根因分析
严重	2025-09-19 18:15:06	应用提供服务统计 接口名称: /api/order/order/getOrdersByCondi...	service-gateway	应用监控	6分钟 3秒	ser 24	通知历史   详情   查看事件   根因分析
严重	2025-09-19 18:14:06	应用提供服务统计 接口名称: /api/order/order/getOrdersByCondi...	service-gateway	应用监控	7分钟 3秒	ser 24	通知历史   详情   查看事件   根因分析
严重	2025-09-19 18:13:06	应用提供服务统计 接口名称: /api/order/order/getOrdersByCondi...	service-gateway	应用监控	6分钟 4秒	ser 24	通知历史   详情   查看事件   根因分析
严重	2025-09-19 18:12:06	应用提供服务统计 接口名称: /api/order/order/getOrdersByCondi...	service-gateway	应用监控	7分钟 4秒	ser 24	通知历史   详情   查看事件   根因分析
警告	2025-09-19 17:48:06	应用提供服务统计 接口名称: /api/order/order/getOrdersByCondi...	service-gateway	应用监控	5分钟 1秒	ser 24	通知历史   详情   查看事件   根因分析
警告	2025-09-19 17:47:06	应用提供服务统计 接口名称: /api/order/order/getOrdersByCondi...	service-gateway	应用监控	6分钟 1秒	ser 24	通知历史   详情   查看事件   根因分析
普通	2025-09-19 16:22:06	应用提供服务统计 接口名称: /api/order/order/getOrdersByCondi...	service-gateway	应用监控	5分钟 6秒	ser 24	通知历史   详情   查看事件   根因分析
普通	2025-09-19 16:21:06	应用提供服务统计 接口名称: /api/order/order/getOrdersByCondi...	service-gateway	应用监控	6分钟 6秒	ser 24	通知历史   详情   查看事件   根因分析

Tooltip content: 应用提供服务统计 接口名称: /api/order/order/getOrdersByCondition 接口调用类型: 0 调用响应时间最近 5 分钟的p99 大于等于 500 毫秒触发普通告警、大于等于 1000 毫秒触发警告告警、大于等于 10000 毫秒触发错误告警、大于等于 20000 毫秒触发紧急告警, 当前值 30028.0000

## Upgrading the AIOps Agent: Now with Natural Language Interaction

The screenshot displays the Alibaba Cloud AIOps console interface. The main area shows a detailed network diagram of a Kubernetes cluster with various nodes and services. On the right, a '智能运维助手' (AIOps Assistant) panel provides a summary of resource usage metrics.

### 内存使用情况 (Memory Usage)

4h

Legend:

- 10.12.19.228 (cn-hongkong.1...)
- 10.12.19.241 (cn-hongkong.10...)
- 10.246.55.11 (cn-hongkong.1...)
- 10.98.248.63 (cn-hongkong.1...)
- 10.246.54.255 (cn-hongkong....)

### 巡检总结 (Inspection Summary)

#### CPU水位分析 (CPU Watermark Analysis)

- 高负载节点: 10.98.248.63 (84.4%), 10.246.55.11 (83.7%)
- 中等负载节点: 10.12.19.241 (21.1%)
- 低负载节点: 其余5个节点CPU使用率均在4-8%之间

#### 内存水位分析 (Memory Watermark Analysis)

- 高内存使用: 10.12.19.241 (51.7%), 10.12.19.228 (50.2%)
- 中等内存使用: 10.246.55.11 (42.4%)
- 低内存使用: 其余5个节点内存使用率在27-41%之间

#### 重点关注 (Key Points)

- 节点 **10.98.248.63** CPU使用率超过84%，需要监控是否存在性能瓶颈
- 节点 **10.12.19.241** 和 **10.12.19.228** 内存使用率超过50%，建议关注内存增长趋势
- 整体集群资源利用率较为健康，无异常检测告警

建议持续监控高负载节点的资源使用趋势，如有业务增长需提前规划资源扩容。

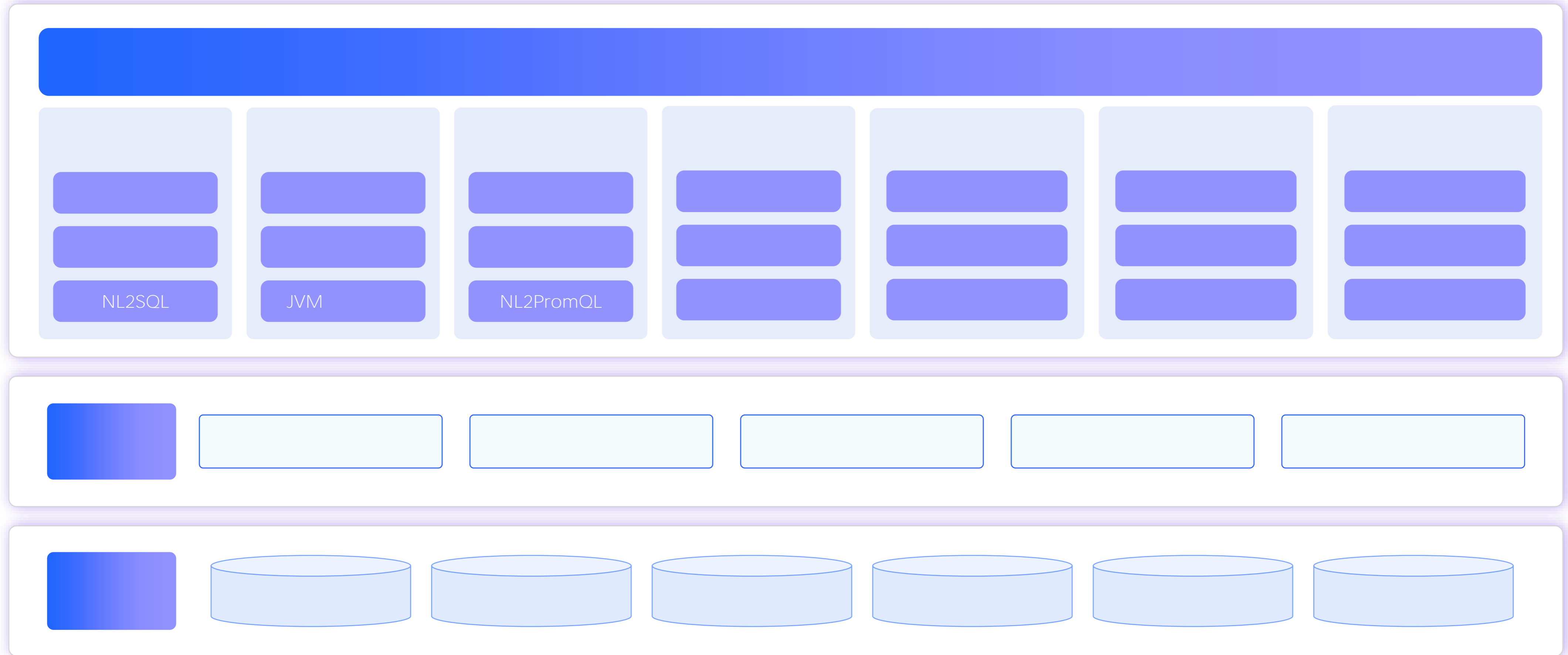
### 智能运维助手 (AIOps Assistant)

请输入你要分析的内容

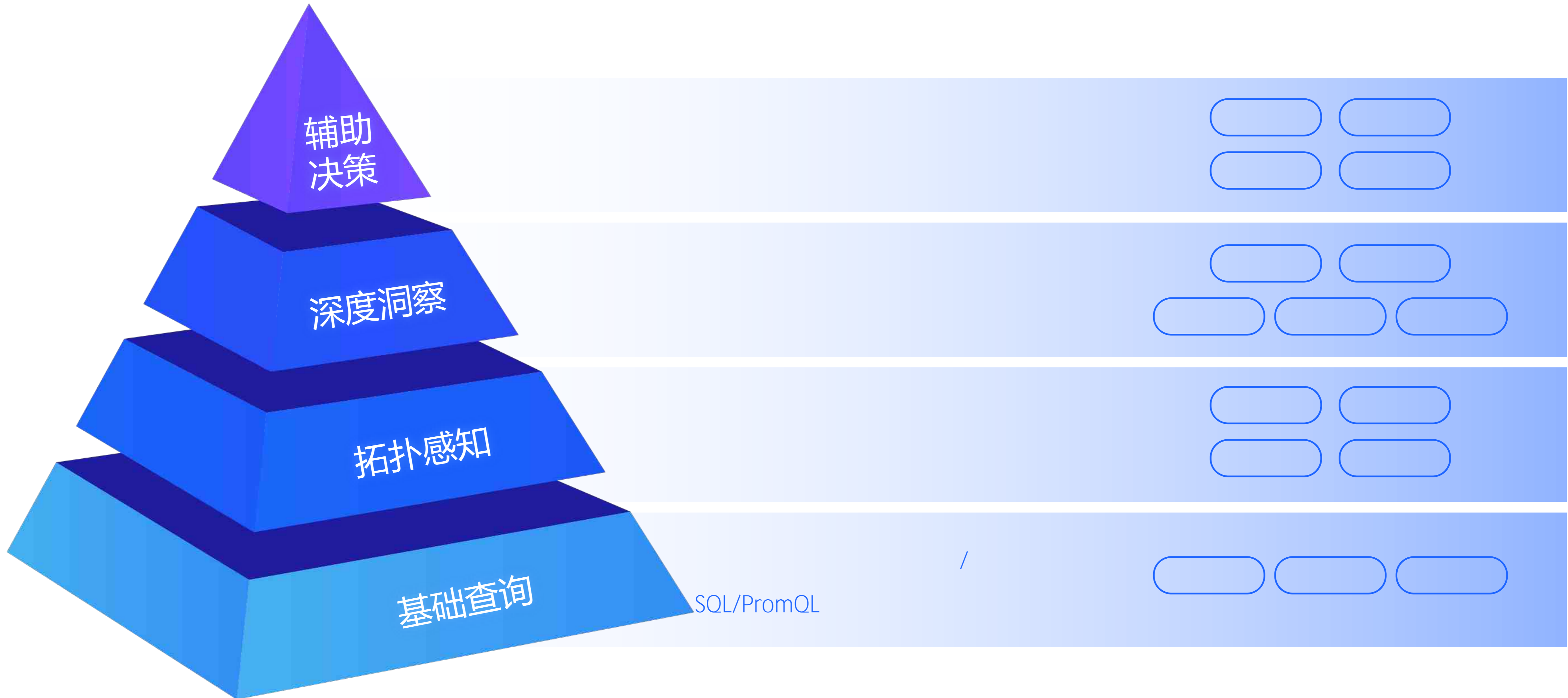


# AIOps

Rebuilding Core AIOps Scenarios with the AIOps Agent



The AIOps Agent: A Layered Capability Stack



The AIOps Agent: Sample Questions



 我接入了多少台 ECS 实例，分别在哪些可用区？

 我有几个 Java 应用？列出他们的 JDK 版本？

 我有哪些实例系列是高可用类型的 RDS？



 应用 xxx 依赖了哪些数据库实例？




 节点 xx 上有几个 Pod 在运行，平均 CPU 是多少？

 这个 K8s 集群上部署的 APM 应用的错误率排名？



 查询 xx 应用上周调用量并做后续三天趋势预测？


 从内存占用维度，检查哪些 Pod 存在异常？




 分析应用的 CPU/内存的性能瓶颈在哪里？



 对 K8s 集群的进行巡检，关注组件与资源水位

 这个应用这个小时有请求超时情况，帮我定位原因

 对比应用发布前后，服务黄金指标以及性能基线

# Demo

Demo: The AIOps Agent in Action

您好，  
欢迎使用

我是您的智能运维助手，可以帮助您分析解决各类场景的运维问题。请注意，我的回答由AI生成，仅作为参考和建议，不能替代您基于专业知识和实际情况的判断。

@ACK:o11y-aiops-demo-cn-hongkong

从 2025-09-17 15:25:30 到 2025-09-17 20:00:00 ，巡检分析这个集群。对遇到的异常资源，进行根因分析，并找到是由什么变更引入的。1. 查询集群中的事件信息，找到异常的事件2. 分析 Node 节点和 Pod 的 CPU 和内存水位是否健康3. 分析 deployments 的变更事件，是否引入了异常风险

🟢 回答完成

我将帮您对集群进行巡检分析，从2025-09-17 15:25:30到20:00:00期间查找异常资源并进行根因分析。

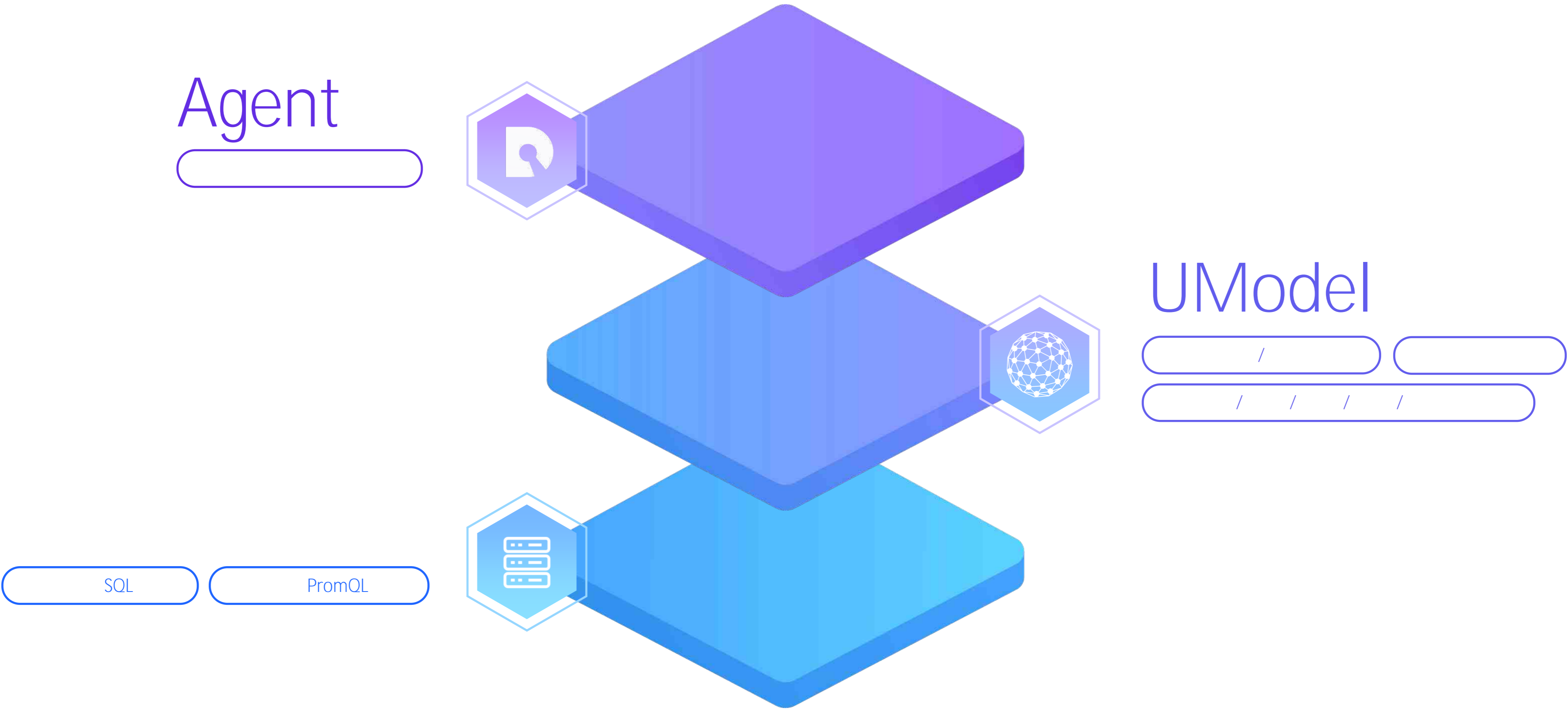
让我先制定详细的分析计划：

任务: 待处理: 6 | 已运行: 0秒

- [ ] 🕒 查询 k8s.cluster(3c6c1d7c86f82c98b007bd377d6f513f) 在 2025-09-17 15:25-20:00 的

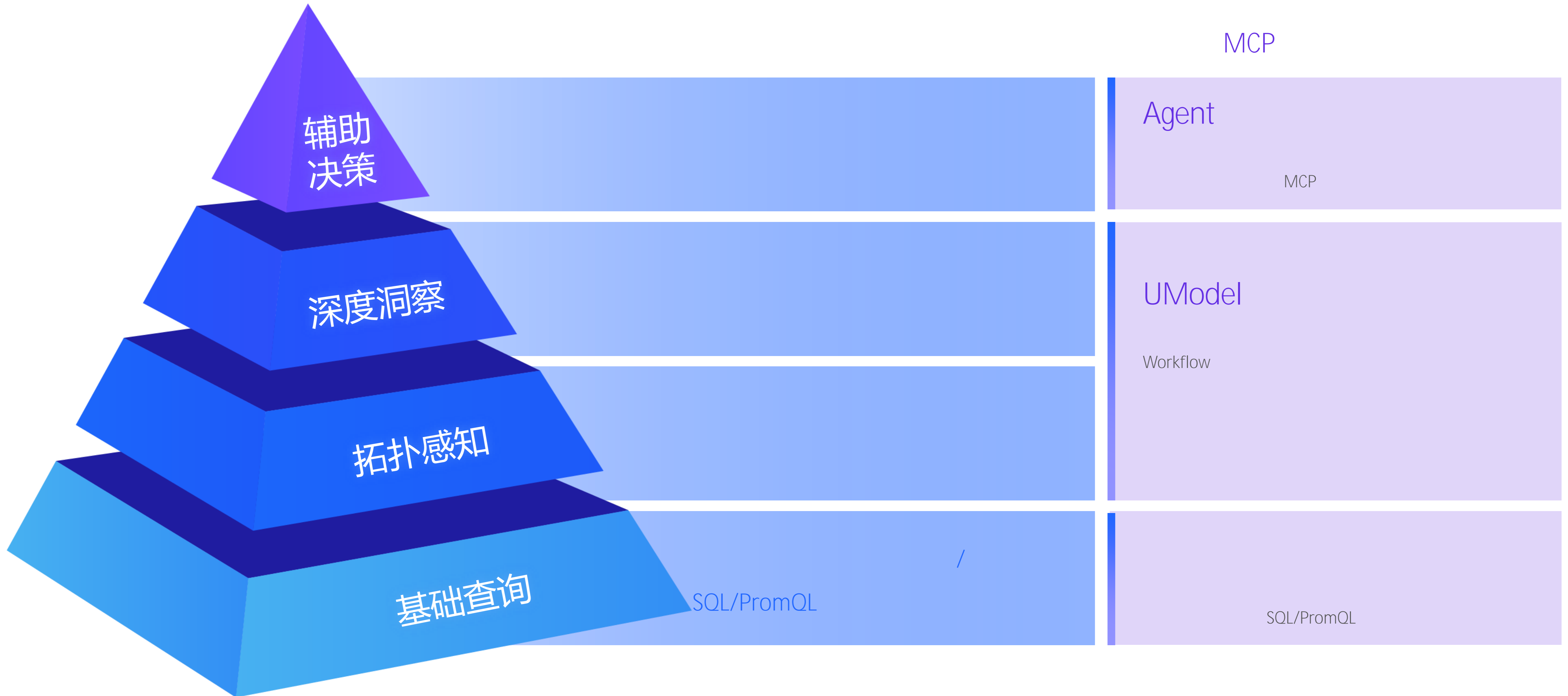
# MCP

Unlocking AIOps Capabilities: The 3-Layer Observability MCP Toolkit



# MCP

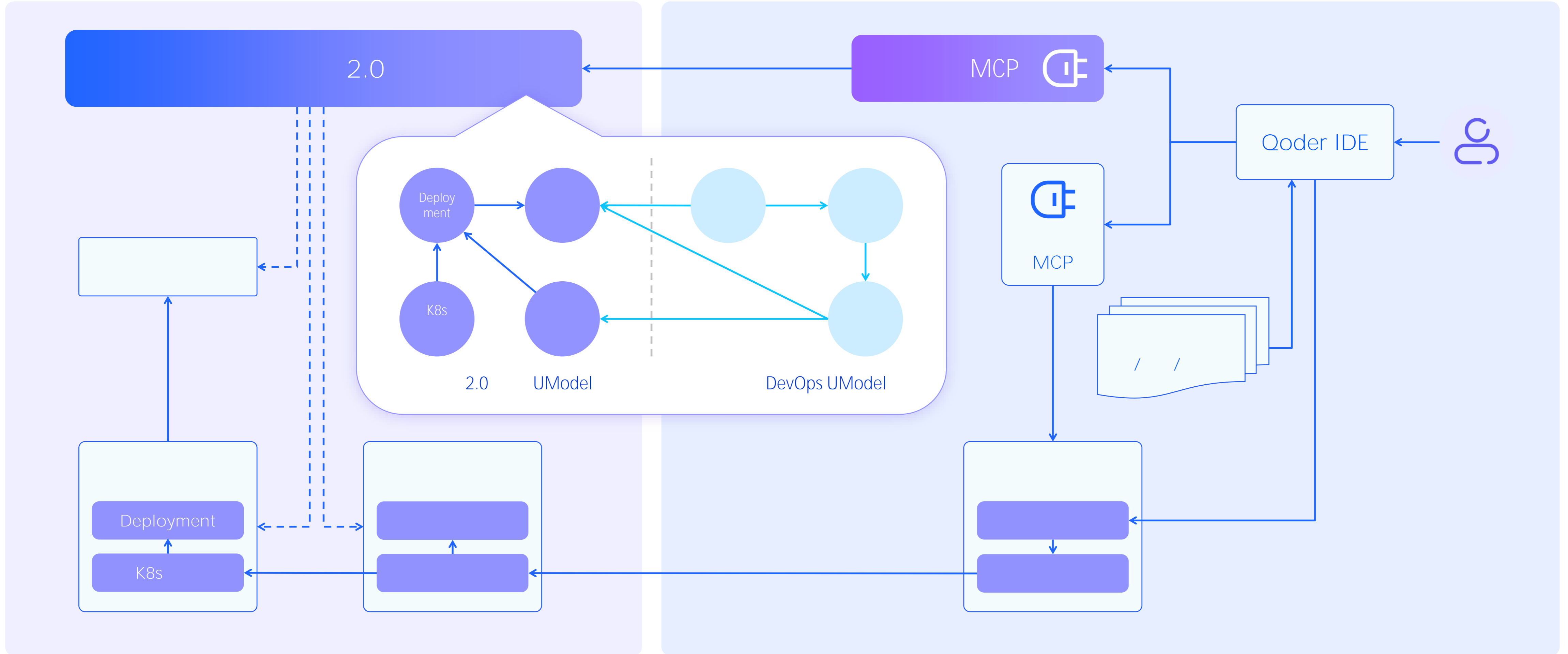
The Observability MCP Toolkit: Key Use Cases



# Demo

# MCP DevOps

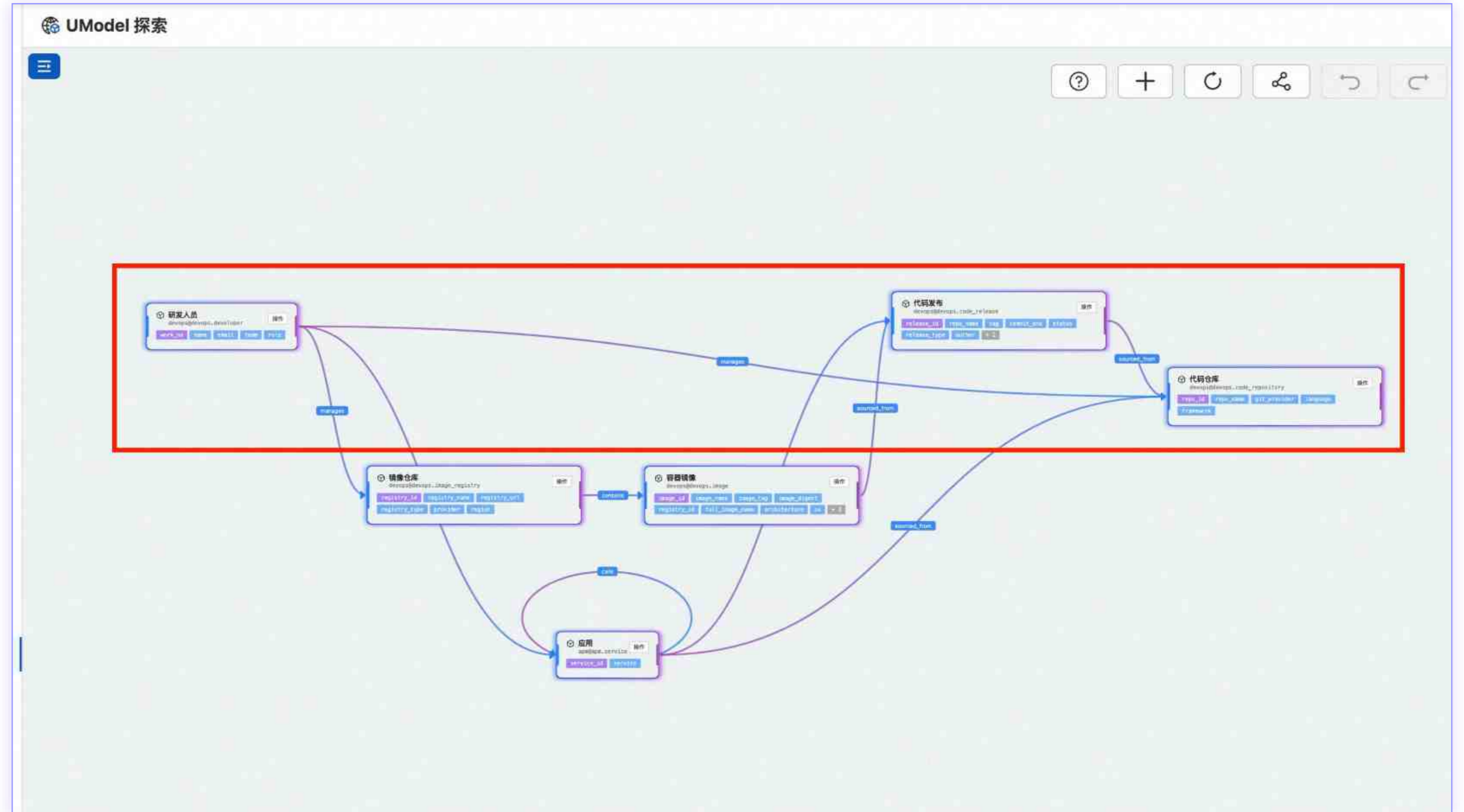
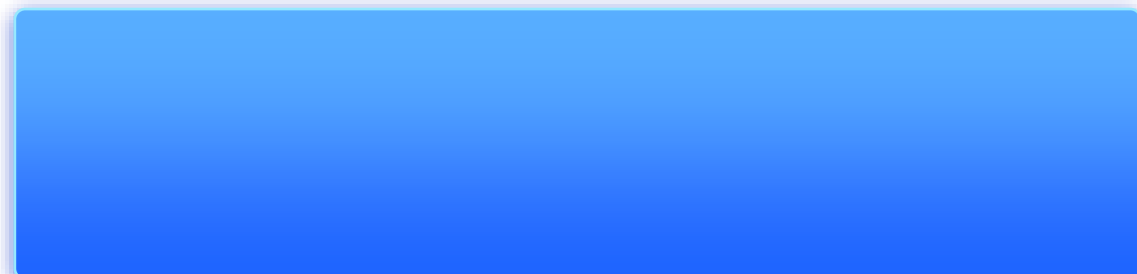
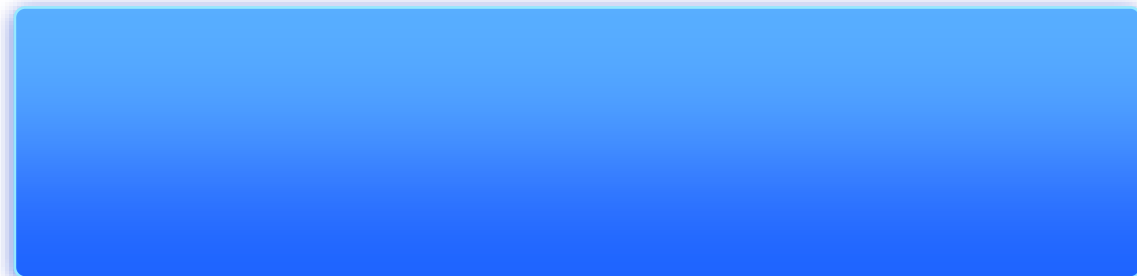
Demo Intro: The Observability MCP in DevOps Scenarios



# Demo

# MCP DevOps

Demo: The Observability MCP in DevOps Scenarios

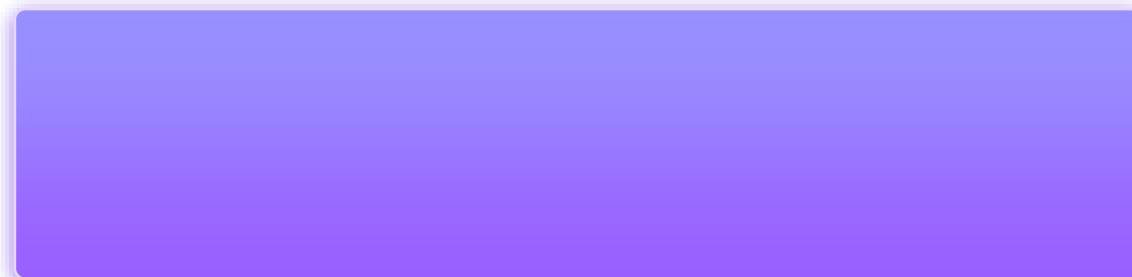
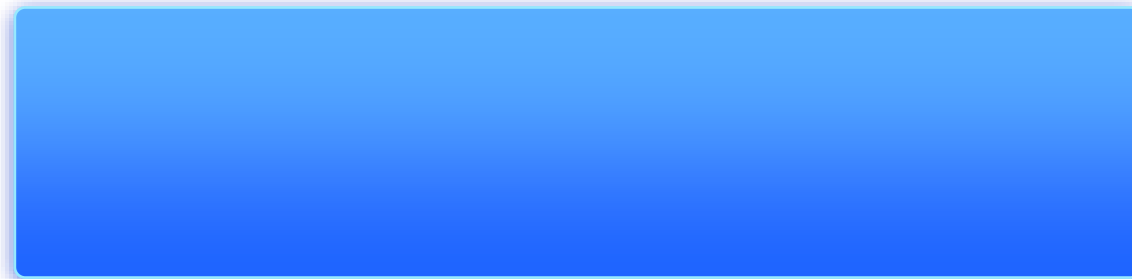
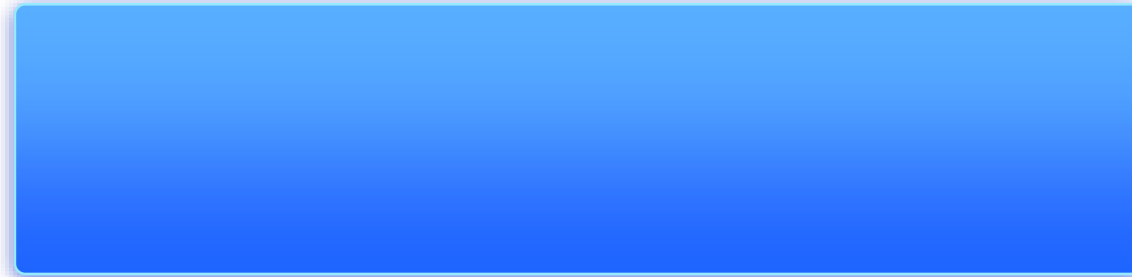


Demo

MCP

DevOps

Demo: The Observability MCP in DevOps Scenarios



+ 添加上下文

我们的服务最近升级的版本出现了问题，请帮我分析代码并找到根因。如果找到了问题，提交Bug并Fix

🔍 根因分析

🔥 直接原因: service-order 镜像升级

变更详情:

变更时间: 2025-09-19 15:46:01

变更内容: mymall-order:v1.0.4 → mymall-order:v1.0.5

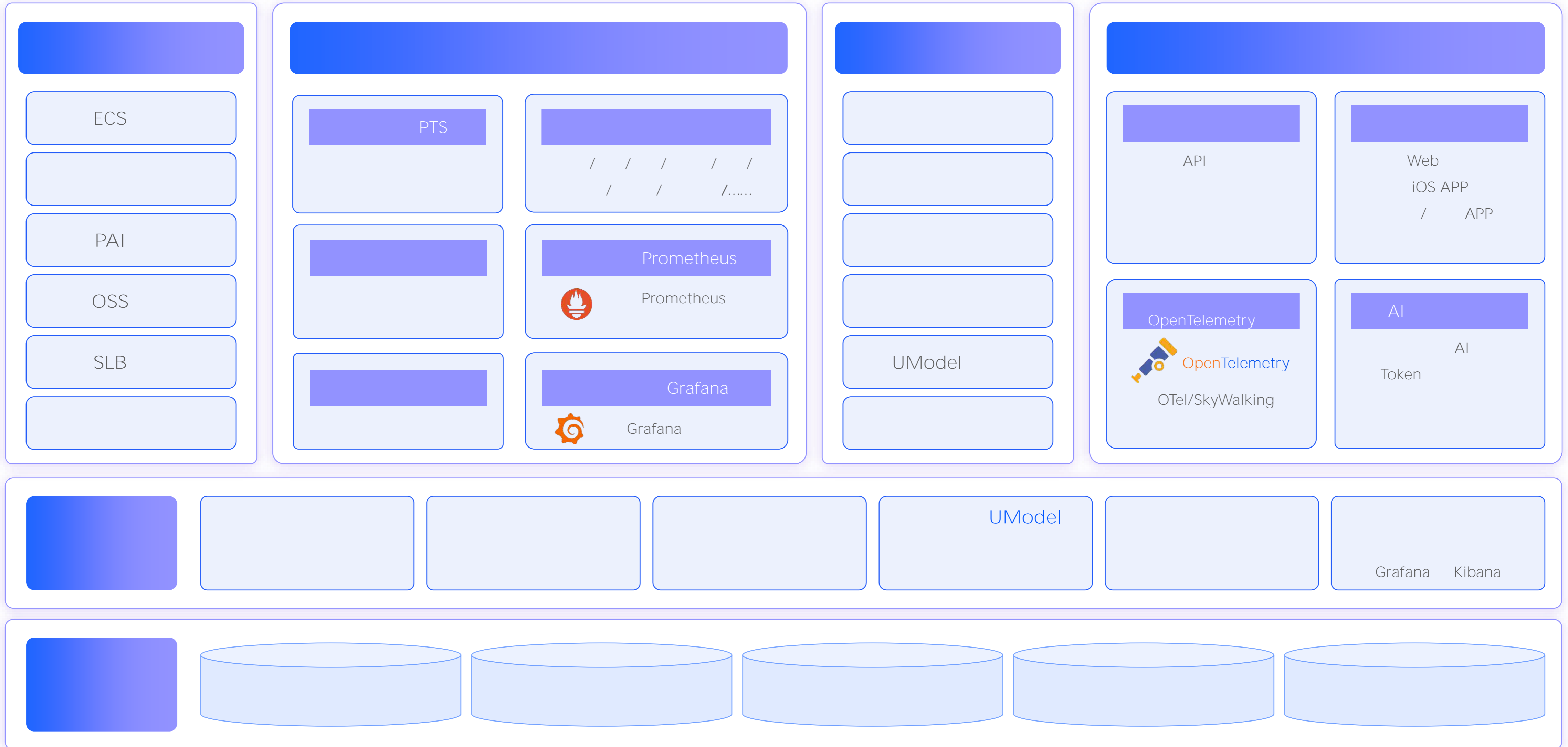
变更作者: yanyadong.yyd


提交信息: 优化订单查询 SQL 条件 - 对 status 字段进行空值处理

智能体 ▾

# 2.0

## CloudMonitor 2.0: Product Landscape



 阿里云 | 计算,为了无法计算的价值

# Apache RocketMQ × AI Agent

2025/10/30

# CONTENT

---

01 Agent AI

---

02 Multi-Agent

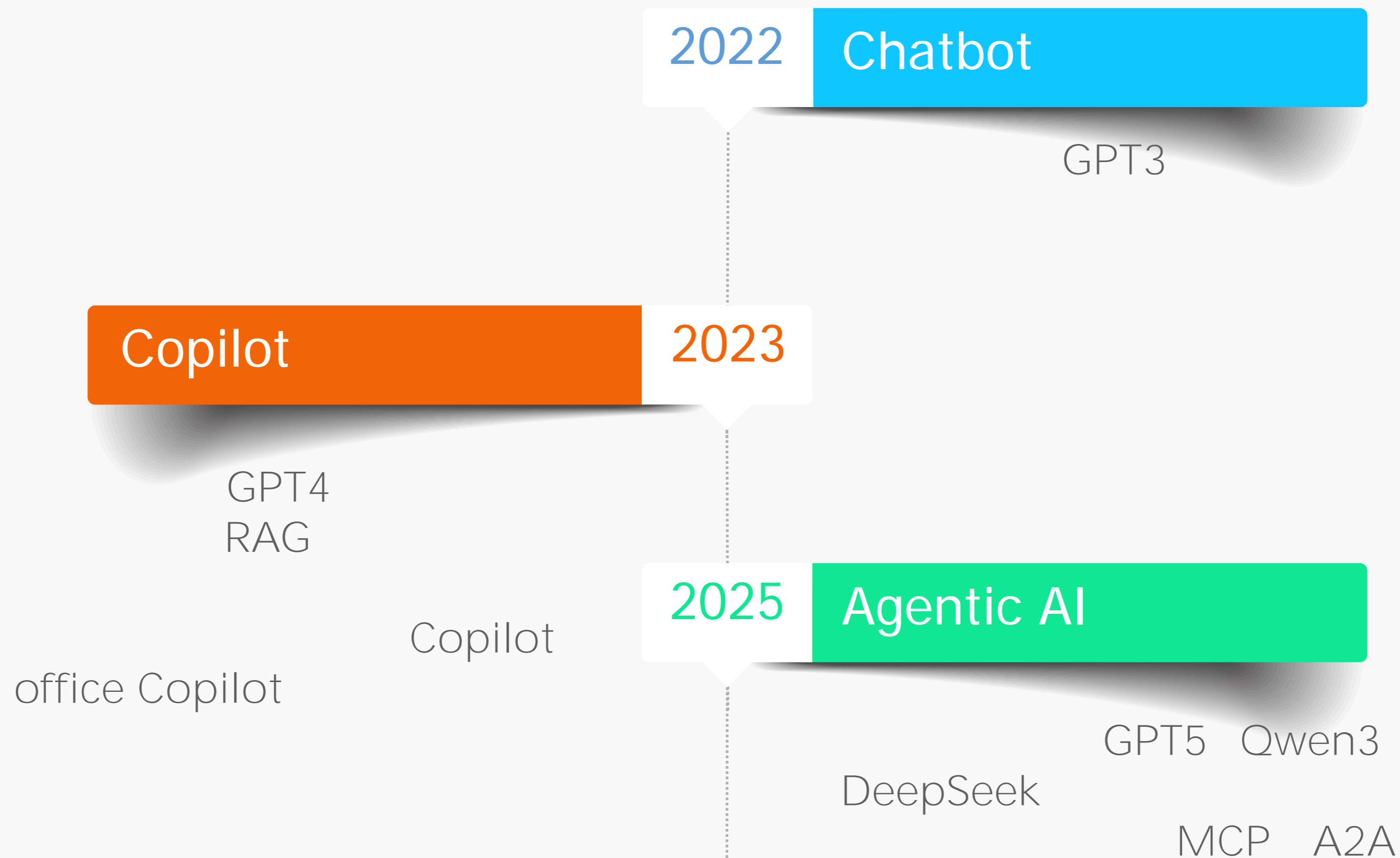
---

03 Apache RocketMQ Multi-Agent

---

# 01 Agentic AI

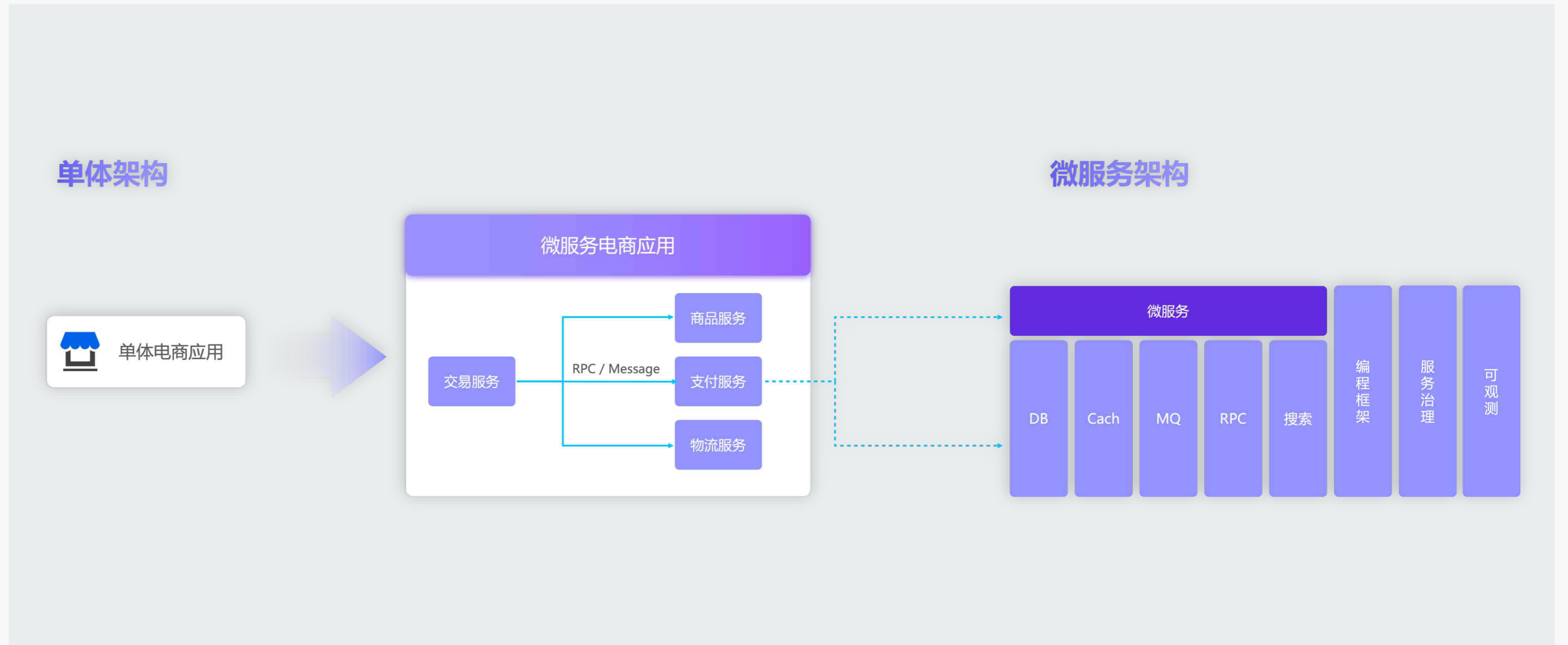
# AI      Agentic AI



# Agentic AI vs

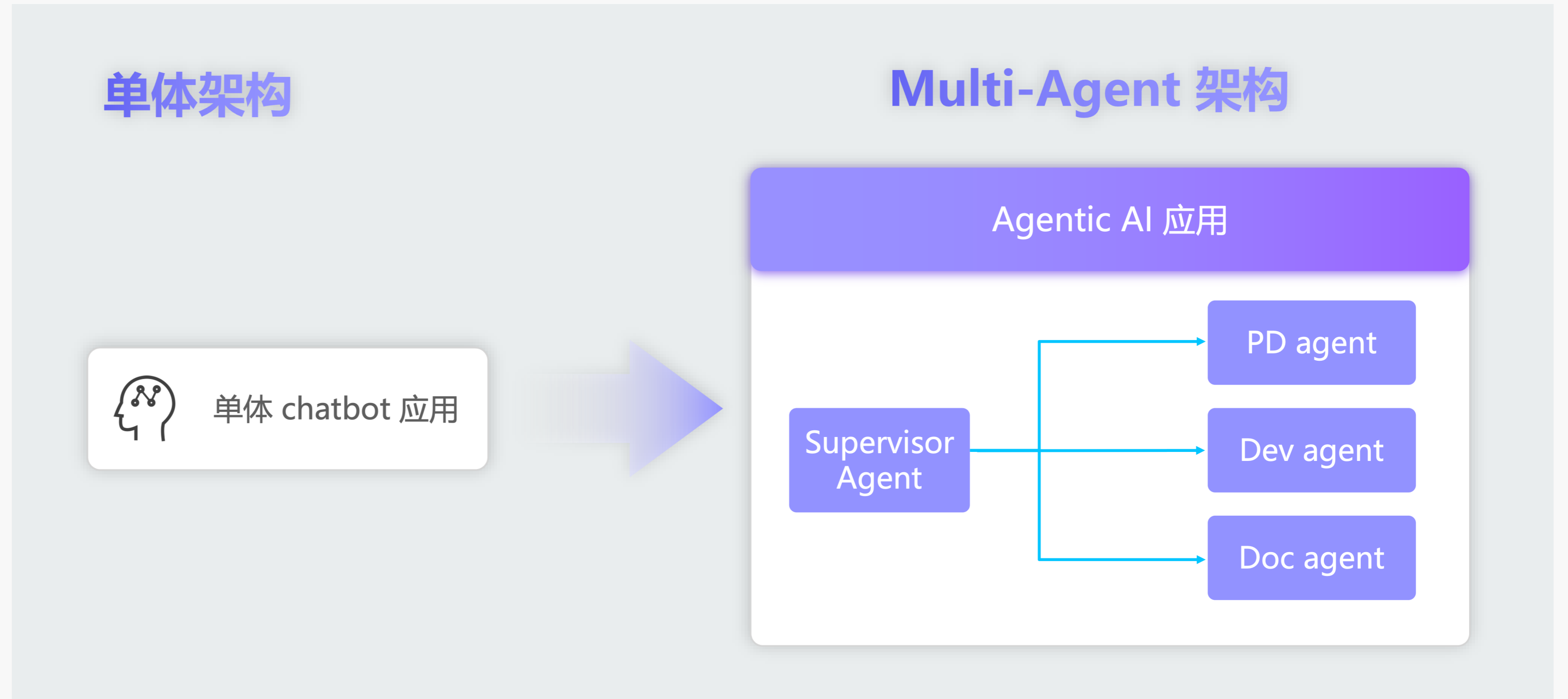
		Agentic AI
	+	" + "
		LLM
		Function Calling / Tool Use
	- API +	+ LLM + +
		+ +
		API +

- 
- 
- 
- 
- 
- 
- 
- 



# Agentic AI vs

- 
- 
- AI
- 
- 
- 
- 



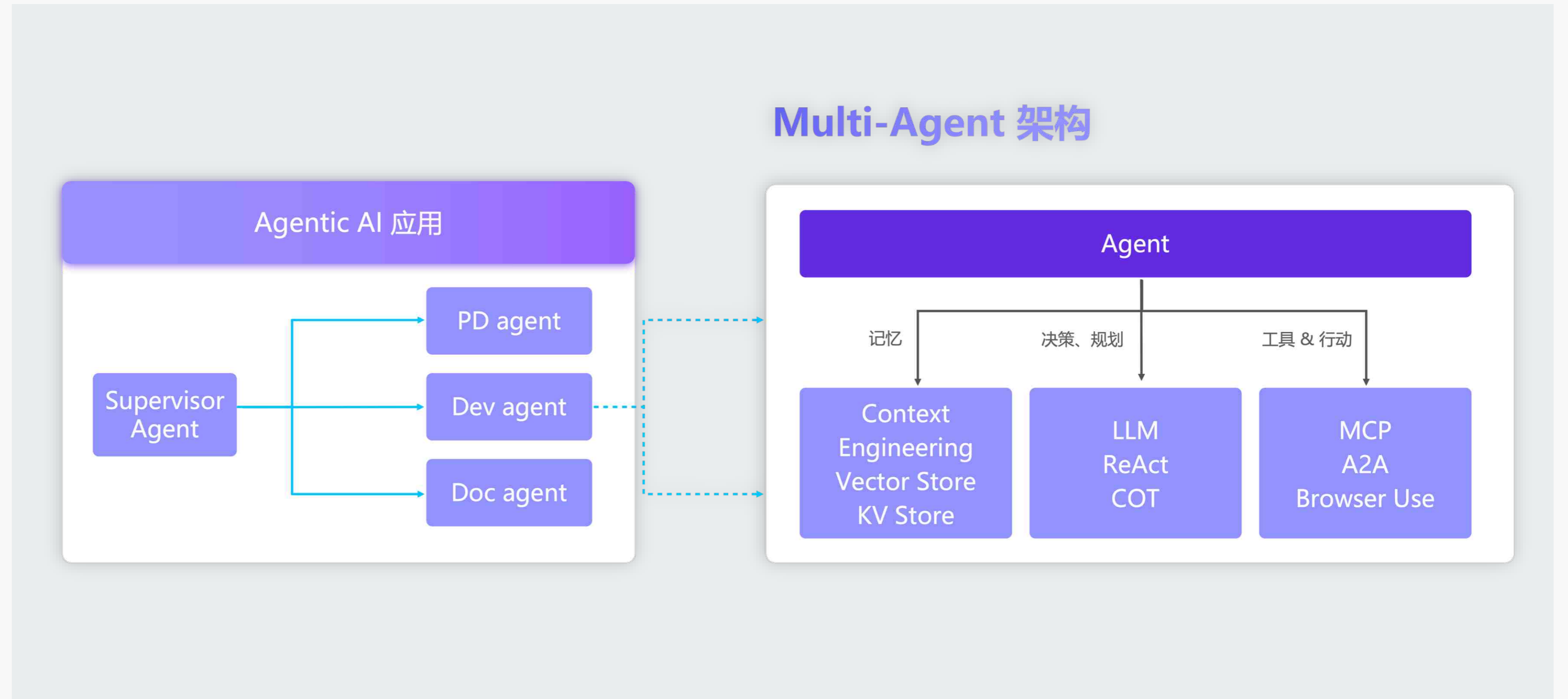
# Agentic AI vs

Supervisor Agent

- 
- 
- 
- 

Sub Agent

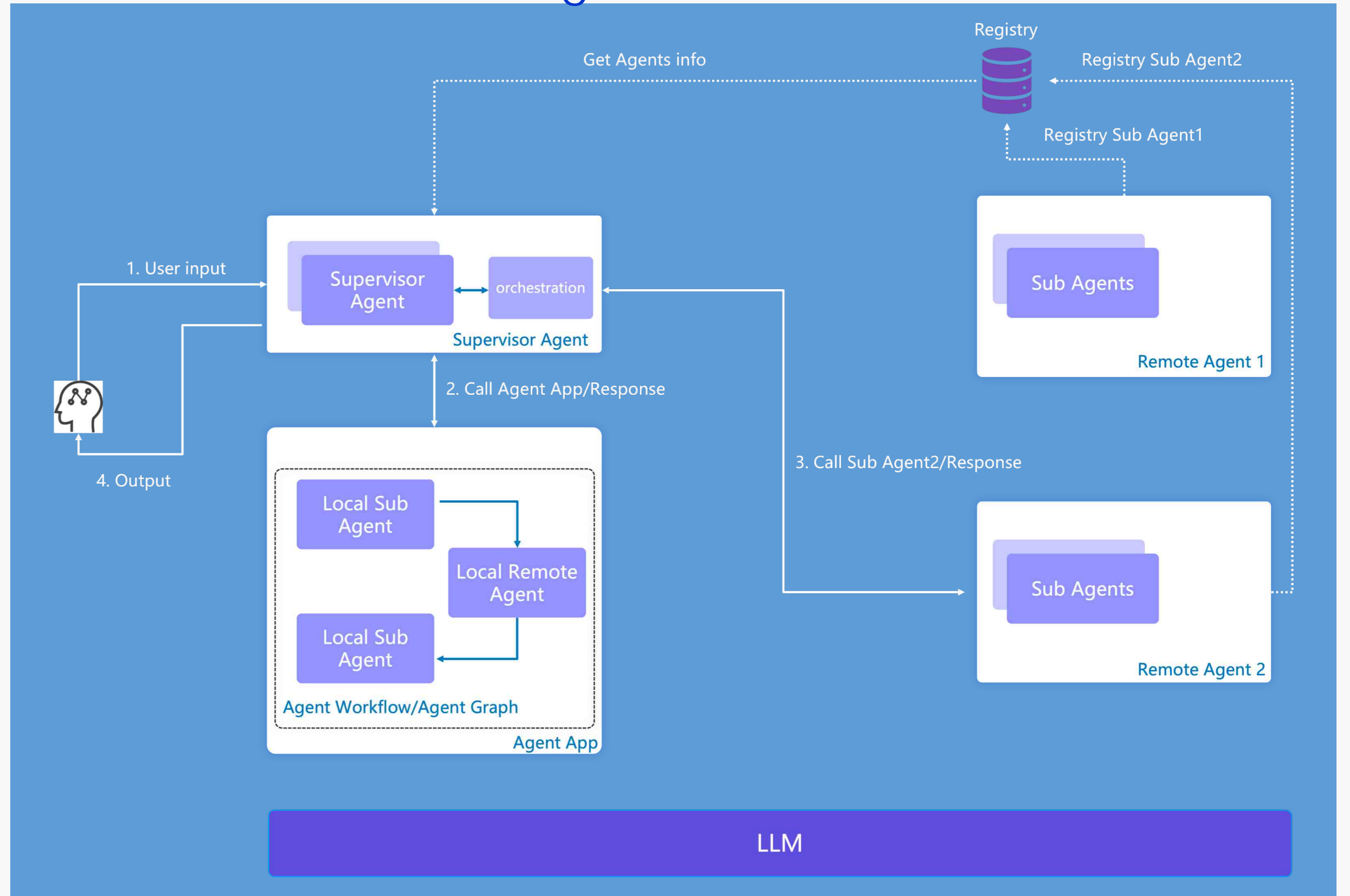
- 
- 
- 
- 
- 



# Multi-Agent

- Sub Agent
- Sub Agent
- Supervisor
- Supervisor
- Agent

Agent

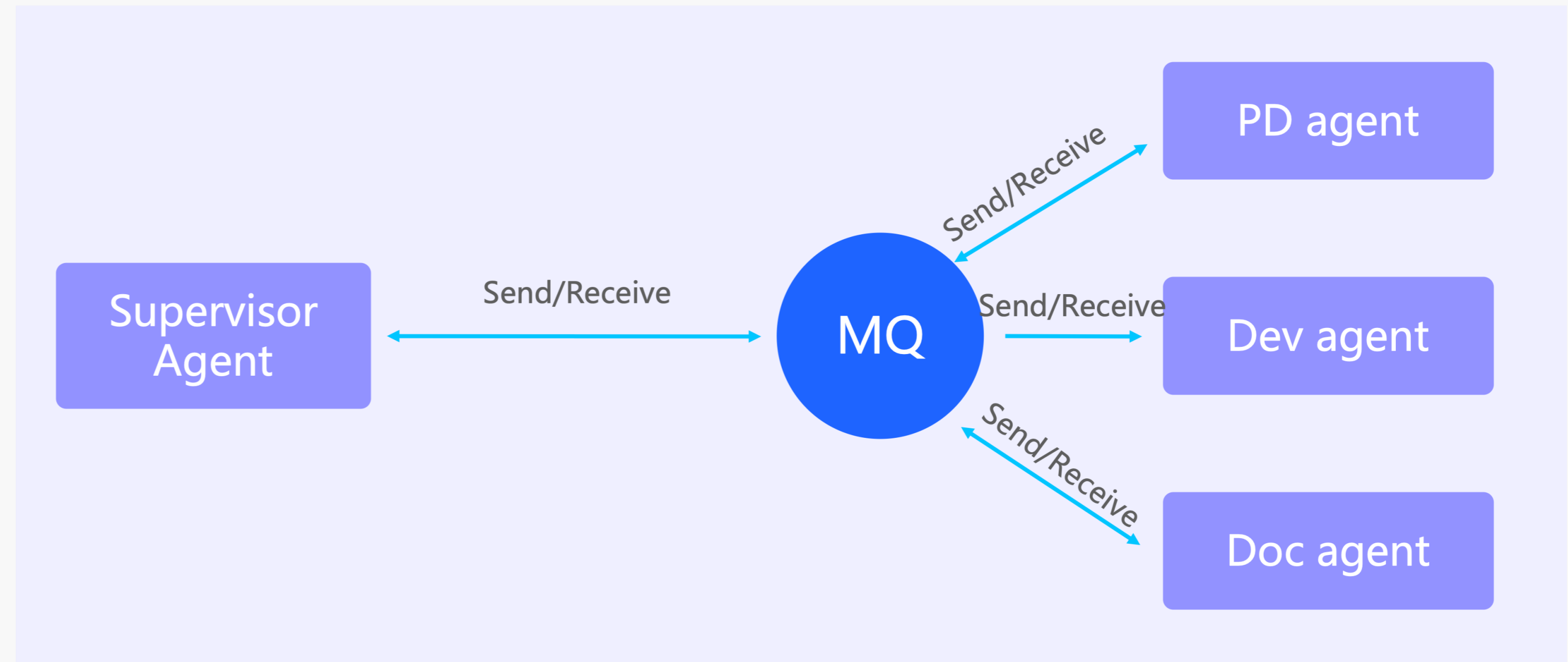


02

Multi-Agent

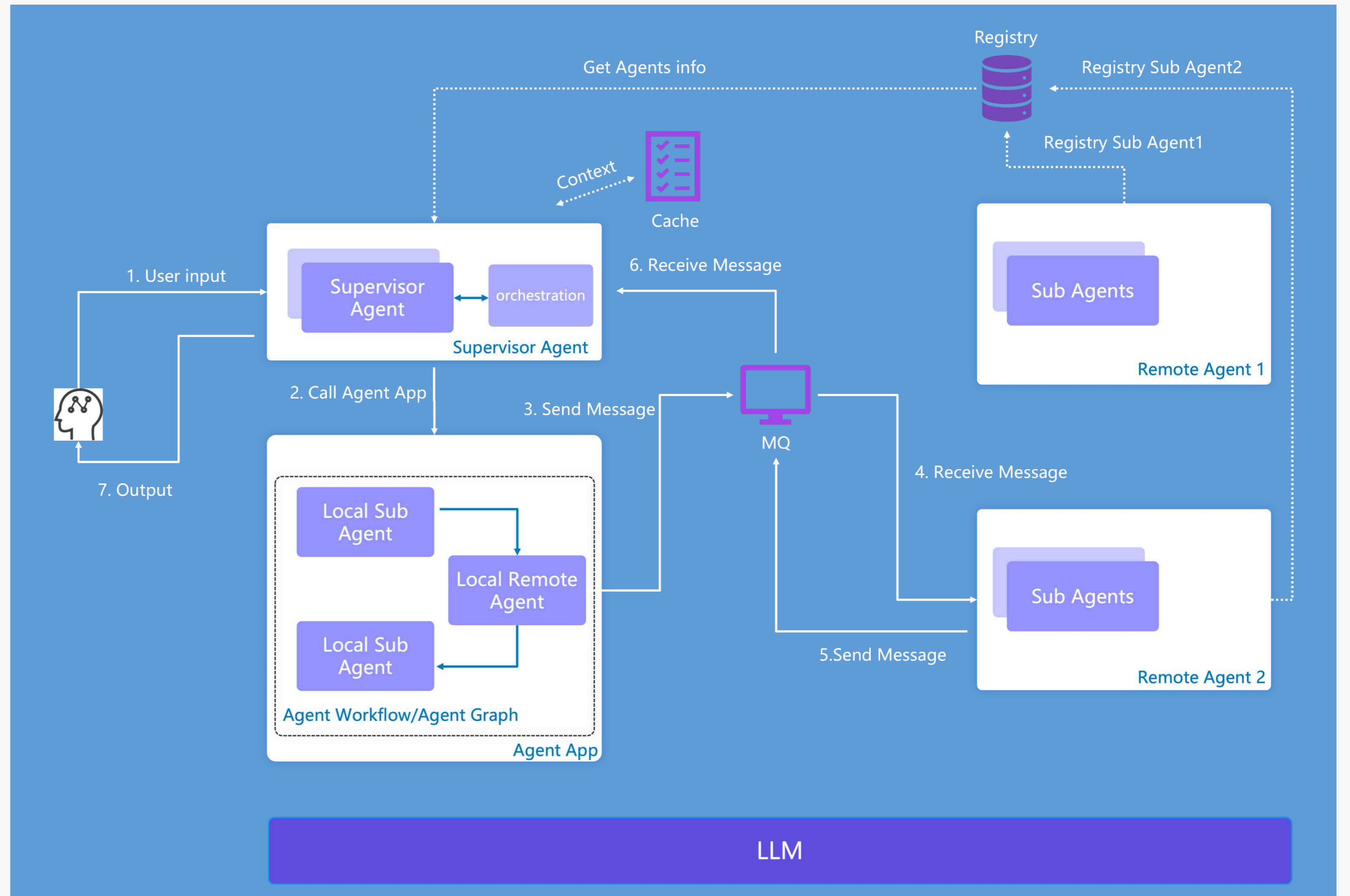
# Multi-Agent

- 
- Supervisor Agent  
LLM
- Supervisor Agent  
Sub Agent
- Sub Agent            Topic



# Agentic AI vs

- Supervisor Agent LLM
- Supervisor Agent Topic MQ
- Topic Agent MQ
- Supervisor Agent Topic
- 



# Multi-Agent

1. Agent Agent Topic

2. Conversation\_Id Task\_Id

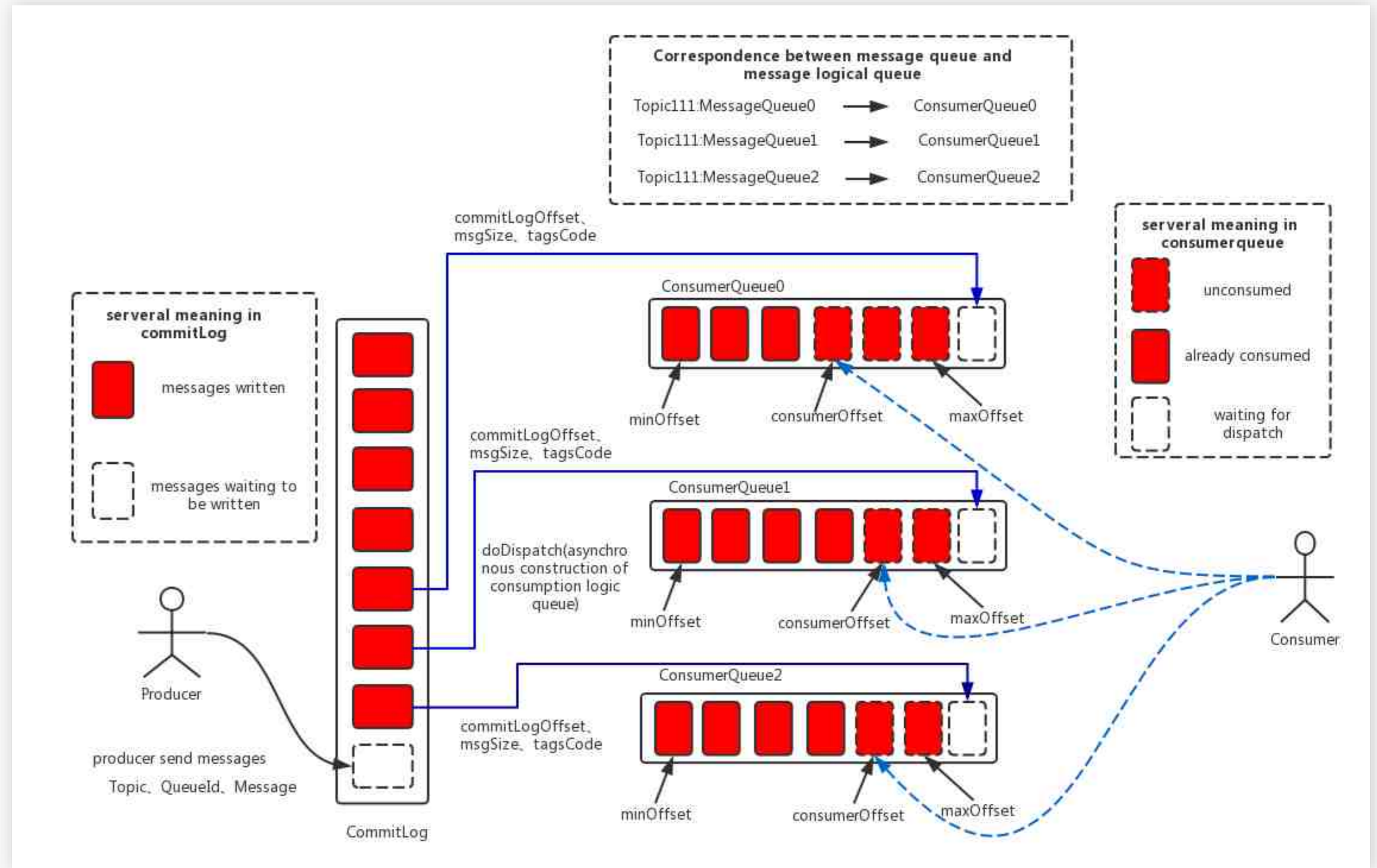
3. Supervisor Agent Sub Agent

Supervisor Agent Sub Agent

# 03 Apache RocketMQ Multi-Agent

# Apache RocketMQ

- (Topic)
- (Queue)
- 
- 
- 



1. Topic  
Agent

2. LLM

3. Supervisor Agent

4. Topic Lite-Topic

# Topic

- Topic Topic
- Topic description  
Nameserver
- Supervisor Agent Topic " "

```

1 {
2   "topic": "query-weather",
3   "queueDatas": [
4     {
5       "brokerName": "broker-a",
6       "readQueueNums": 8,
7       "writeQueueNums": 8,
8       "perm": 6,
9       "topicSynFlag": 0
10    }
11  ],
12  "brokerDatas": [
13    {
14      "brokerName": "broker-a",
15      "cluster": "DefaultCluster",
16      "brokerAddrs": {
17        "0": "192.168.1.10:10911",
18        "1": "192.168.1.11:10911"
19      }
20    }
21  ]
22  "appCard": {
23    "name": "WeatherQueryAgent",
24    "description": "Query current weather by city name.",
25    "capabilities": [
26      {
27        "intent": "get_weather",
28        "description": "Get current weather for a city",
29        "input_contract": {
30          "type": "message_properties",
31          "required_properties": ["city_name"]
32        }
33      }

```

# Lite-Topic

1. Lite-Topic

2.

3. Lite-Topic

4.

1. Topic

2.

3. Topic Session\_id User\_id

# Lite-Topic vs Topic

	Topic	Lite-Topic
topic		
Topic		
Topic		

# Lite-Topic AI

- Producer

LLM

- Agent

Cache

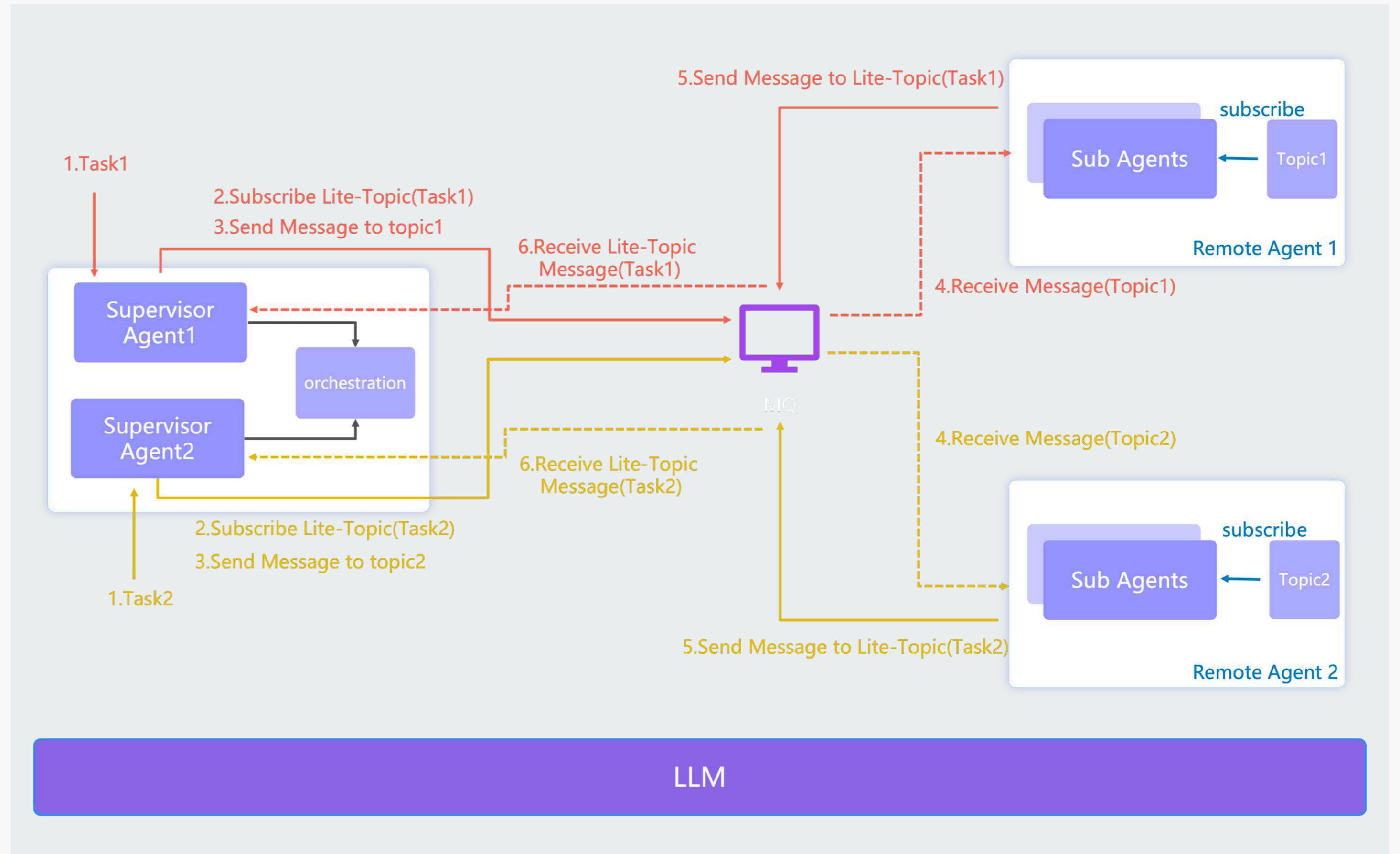
- Lite-Topic

Multi-Agent  
Sub Agent

Supervisor Agent

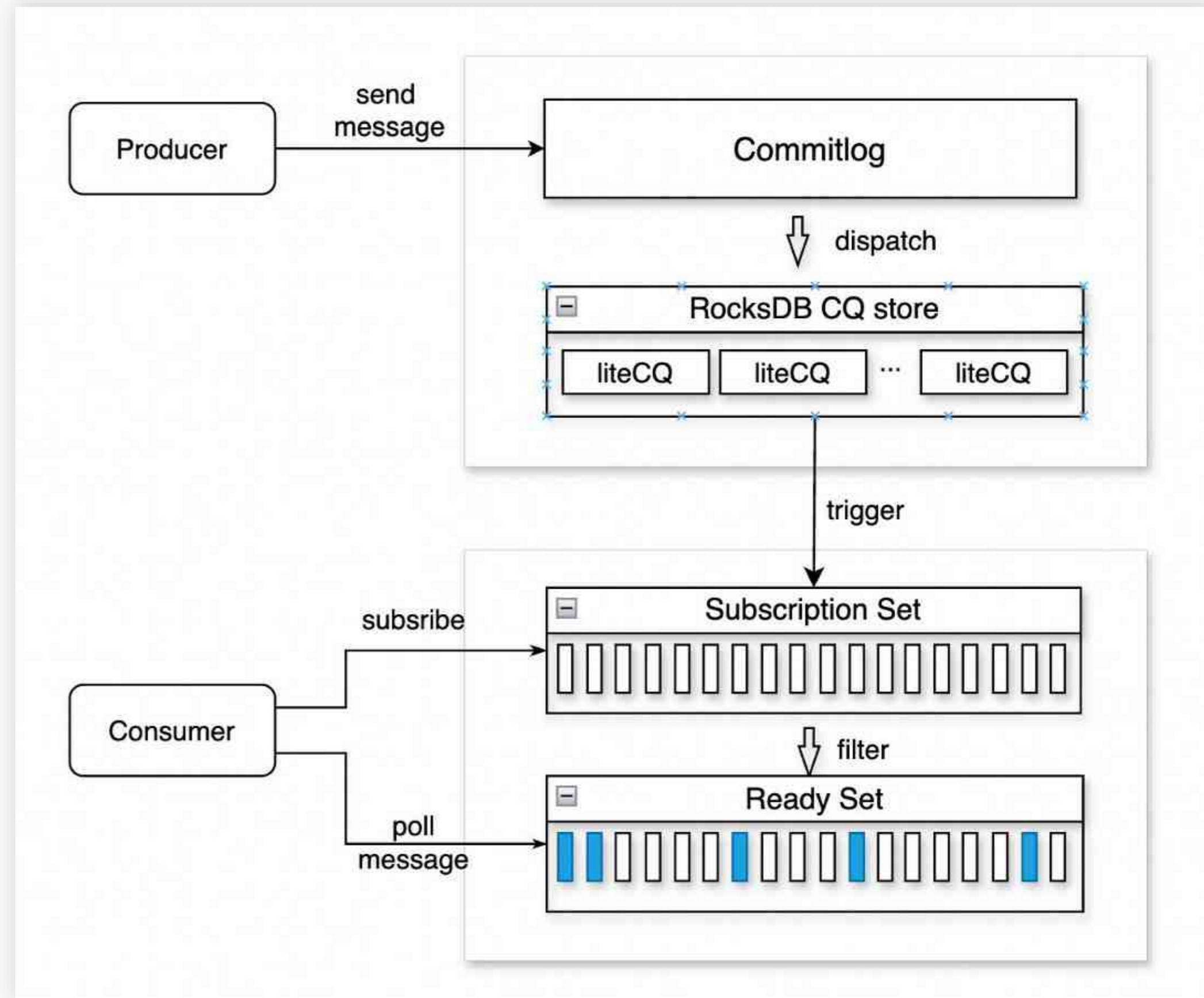
Supervisor

DB



# Lite-Topic

- Lite-Topic RocksDB
- Broker Subscription Lite-Topic Set
- Broker Subscription Set
- Ready Set
- Ready Set Broker poll

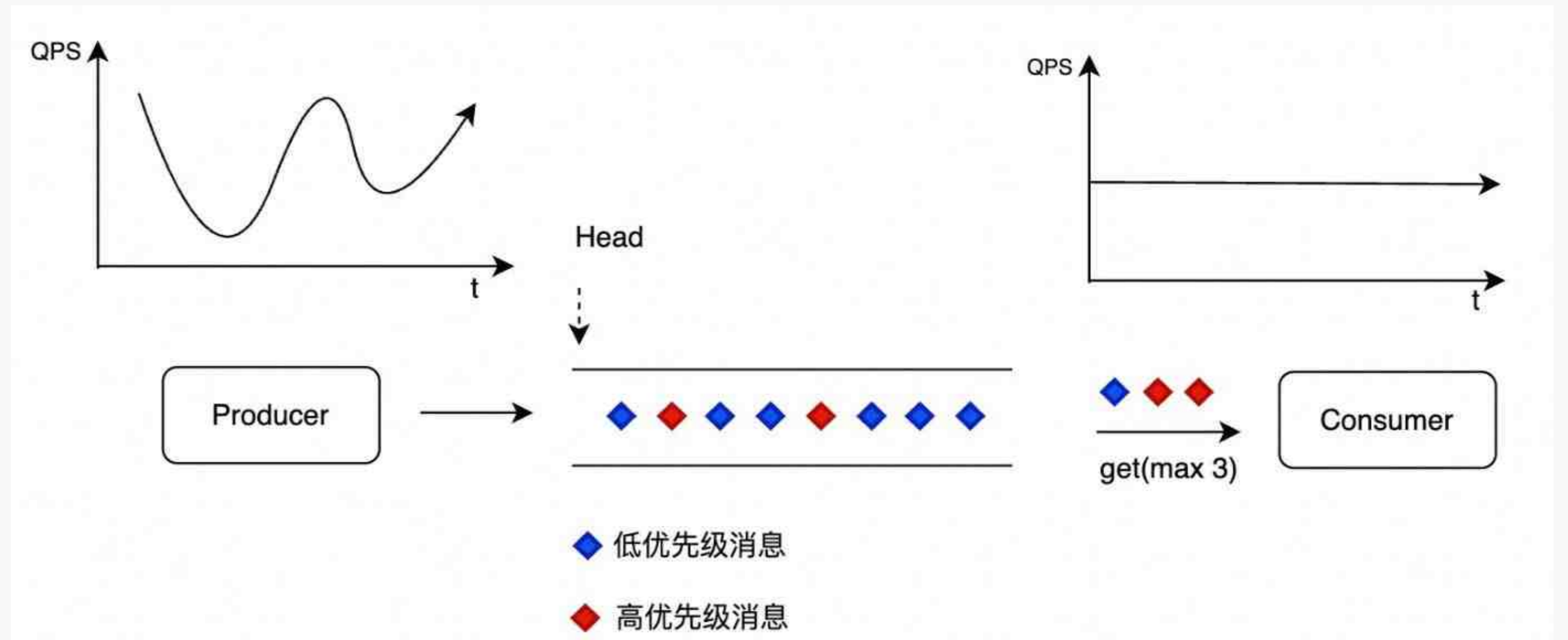


- 
- 
- 
- 
- 
- Apache RocketMQ

“ ”

AI

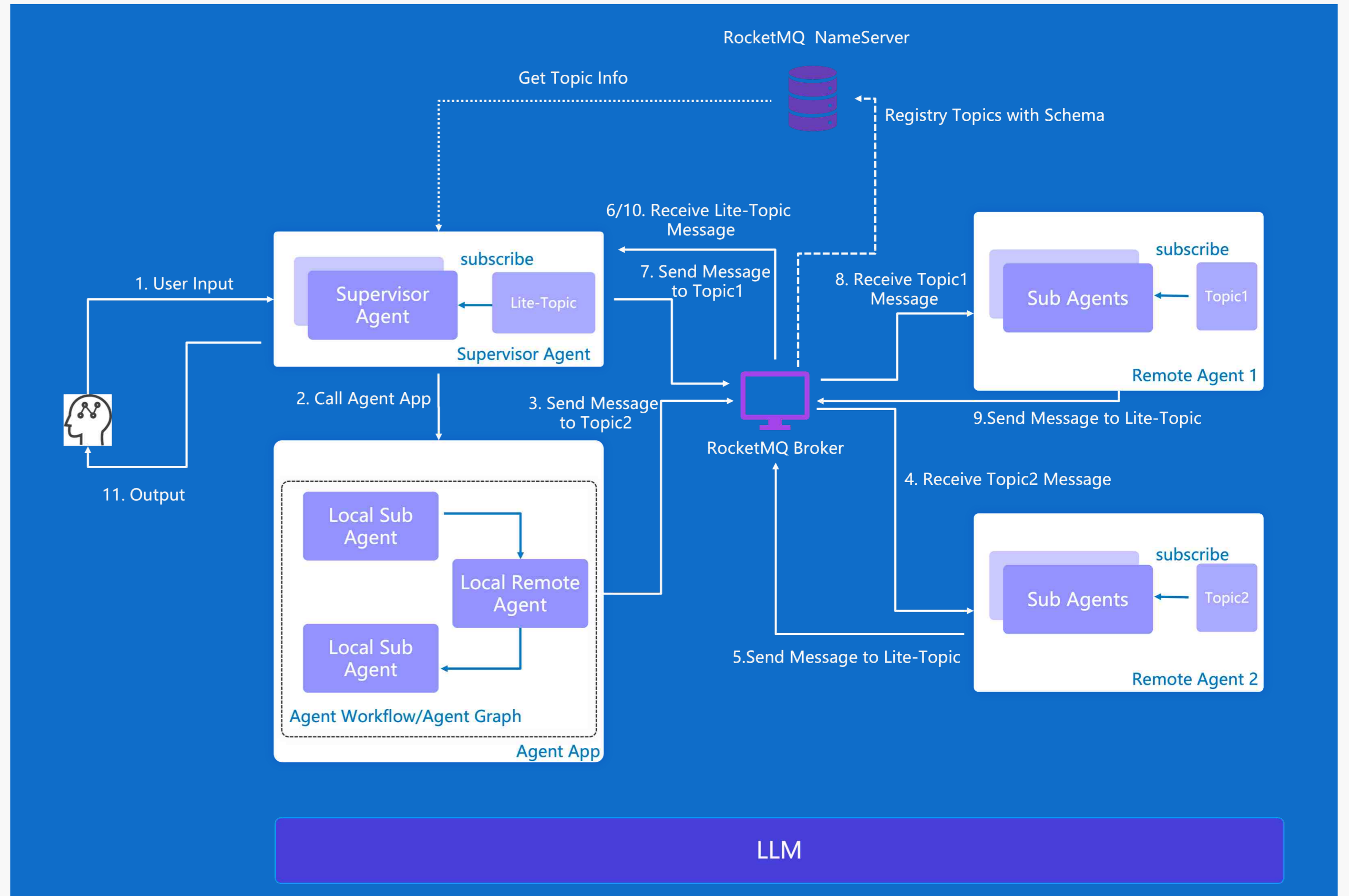
AI



# RocketMQ Multi-Agent

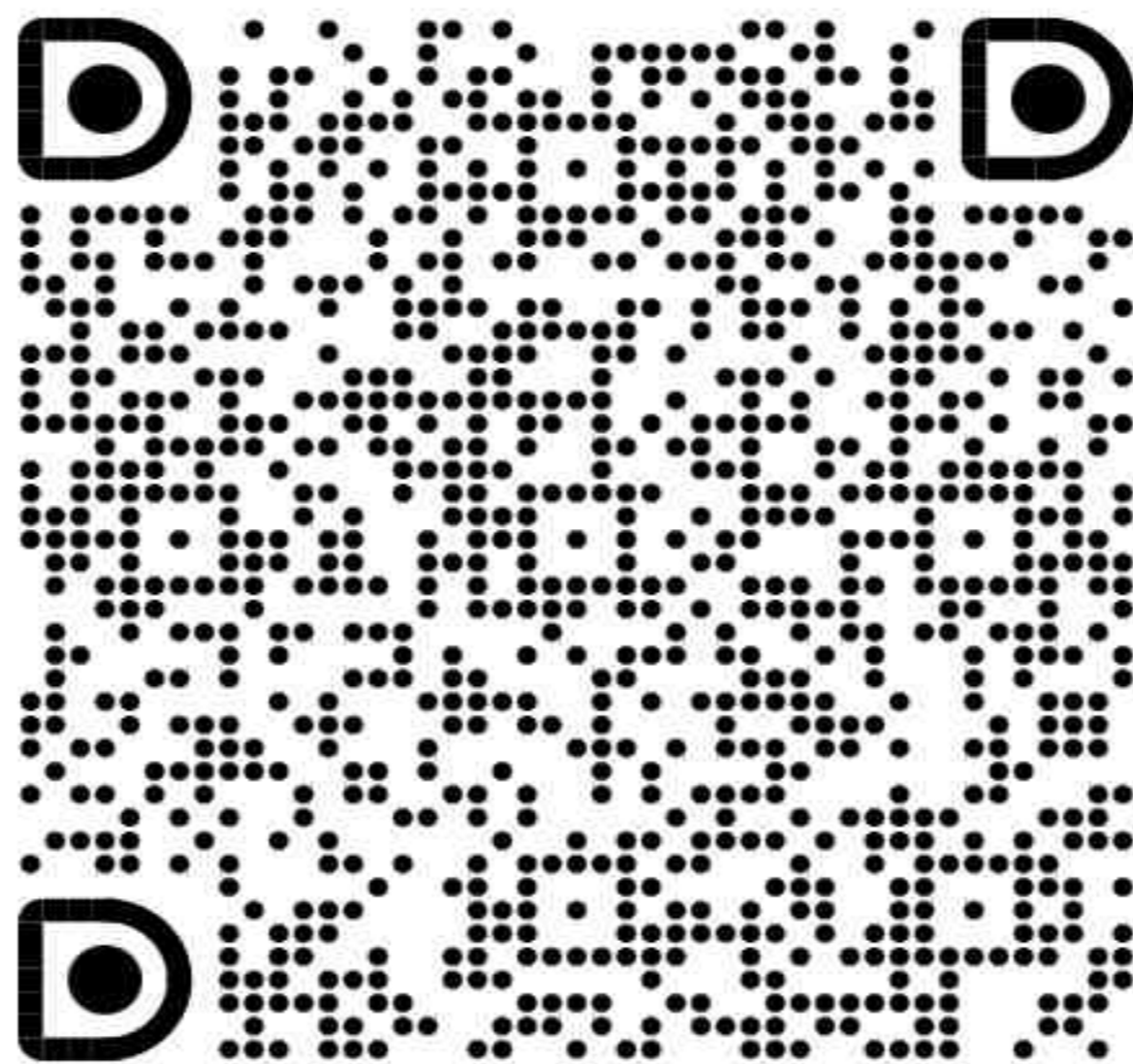
- Sub Agent  
Topic description  
NameServer
- Supervisor Agent  
Prompt
- LLM
- Supervisor Agent  
Topic  
Lite-Topic
- Supervisor Agent

Topic




RocketMQ for AI 用户交流群 普通群

38人



此二维码365天内有效 (2026-10-10 前)

 钉钉扫一扫群二维码，立即加入群聊

 阿里云 | 计算,为了无法计算的价值