川序

DATABASE INFRASTRUCTURE FOR AI AGENTS

川序白皮书

面向企业 AI Agent 的数据库基础设施与协作治理

公开版 · 中文

产品基线：v5.0.4

软件打包日期：2026 年 11 月 5 日

文档日期：2026 年 11 月 5 日

目录

01 阅读说明

02 执行摘要

03 1. 从单次问答到持续业务

04 2. 产品定位与设计原则

05 3. 总体架构与权威数据模型

06 4. 人、Agent、组织与安全域

07 5. 数据库原生多信号检索

08 6. 有来源的知识入库与发布

09 7. 时态记忆、工作便笺与上下文

10 8. Task、Graph、Loop 与受控执行

11 9. Work、Handoff 与 DB4A2A 协作

12 10. 工具、Skill、协议与框架接入

13 11. 模型路由、容量、用量与费用

14 12. 安全、治理与合规职责

15 13. 数据库适配与部署条件

16 14. 安装、升级、运维与恢复

17 15. 版本、许可与学习资料

18 16. 应用场景与工作方式

19 17. 验证方法与当前证据边界

20 18. 概念验证、选型与实施建议

21 附录 A：术语表

22 附录 B：资料依据与继续阅读

阅读说明

本白皮书面向企业技术负责人、架构师、数据库管理员、信息安全与合规团队，以及负责 AI 应用和 Agent 交付的实施团队。它说明川序的产品定位、技术架构、主要业务链路、部署条件和评估方法，便于读者判断是否适合自己的环境，并据此设计概念验证。

本文以 v5.0.4 的当前产品、交付说明与验证记录为依据。文中“支持”均受所选版本、数据库、授权、配置及部署前提约束；协议适配只覆盖明确列出的版本与操作。示例场景用于解释设计，不代表已落地客户案例。建议的实施步骤不构成交付周期、性能、合规认证或服务等级承诺。采购和生产使用以相应软件许可证及合同为准。

中英文版采用相同章节与事实口径。产品升级后，应以实际交付包中的版本说明、部署指南、许可证和迁移说明确认差异。附录 B 列出本文依赖的产品资料；本文不能替代安装手册或数据库安全配置指南。

执行摘要

川序（Chuanxu）是以数据库为权威状态源的企业 AI Agent 基础设施。它把人员与 Agent 身份、组织与安全域、知识与记忆、任务与协作、模型与工具调用，以及审批和执行证据连接起来。企业可以在自己的环境中部署，并选择 Oracle、PostgreSQL 或 YashanDB 作为数据底座。[R1]

川序关注的是 Agent 持续参与业务之后的问题：谁对它负责，它现在可以读取什么，使用的是哪份知识，任务停在哪里，交接是否被接受，外部操作到底有没有完成，以及权限撤销后旧上下文还能否使用。这些问题需要可核查的状态、版本和授权，不能只依赖对话记录或模型对任务状态的描述。

其技术路线有三个重点。第一，业务实体与执行记录保存在数据库中，图关系和摘要保留原始来源，不能成为另一套授权依据。第二，在读取、准备上下文、发送模型请求、调用工具及披露结果时检查当前权限；协作邀请、文档引用和 Agent 自述都不能授予资源权限。第三，把计划、批准、派发、实际执行和结果验收分别记录；未知结果保留为未知，再通过证据恢复。

v5.0.3 贯通了关系属性、向量、全文、标签与有界图关系在知识问答和任务上下文中的使用，并提供有来源、适用范围和责任人的引导入库。排序采用数据库查询与确定性融合，不增加生成模型调用；查询向量、上下文和回答生成仍有各自成本。检索命中也不等于证据充分或答案正确。[R2–R4]

公开产品包括一个可选三种数据库的社区版，以及分别面向 Oracle、PostgreSQL 和 YashanDB 的三个企业版。社区版提供公共的知识、记忆和协作基础；企业版增加专属审批、审计管理、LDAP、Skill Token、编排治理及合规模块。企业版的生产使用需要商业许可。选择产品时，应同时核对业务治理要求、数据库条件和实际模型能力，而非只比较界面功能。[R15]

v5.0.4 增加官方事件来源接入、带证据的故障复盘草稿和上下文装配依据，并加强最终模型输入与完整输出检查。公开事件可用于说明供应商侧事实，不能代替企业内部影响、未知根因和处置验收。准备成本测量没有启用 query/TTL 缓存；共享编码仍是关闭的研究项。[R20, R22]

1. 从单次问答到持续业务

单次问答可以在一个请求中完成。持续业务则可能跨越多个工作日、人员、Agent 和外部系统。一个技术支持任务可能先查手册，再整理诊断证据，交由另一名 Agent 复核，最后由人批准变更。任何一次权限变化、文档修订、网络中断或任务重试，都可能改变后续动作能否继续。

如果这些状态散落在聊天窗口、向量库、脚本日志和框架内存中，企业很难回答几个基础问题：输出依据哪一版材料，某次读取是否仍然合法，旧执行器是否还在写结果，以及恢复任务是否会重复产生外部副作用。模型即使能给出流畅解释，也不能替代这些记录。

川序将上述问题组织为同一条业务链：人设定目标和授权，Agent 发现并读取允许使用的知识，Task 与 Step 记录工作，工具或模型执行受控动作，Work 与 Handoff 承载责任和交接，结果依据真实回执进入后续流程。需要长期复用的经验再进入独立审核和发布，而不因一次回答就自动成为企业知识。

川序适用于需要持久状态、多个主体协作、来源追溯与权限治理的 Agent 应用。若需求仅为一次文本生成，直接模型接口可能已经足够。已有应用开发平台或 Agent 框架可以在明确的协议和执行边界内接入川序；接入不意味着所有外部框架能力均被托管或验证。

2. 产品定位与设计原则

川序提供数据、执行与治理的共同基础。它可以服务知识问答，也可以承载长任务、图执行、有限循环及跨 Agent 交接。数据库提供持久状态和原生权限机制，模型负责在获准输入上生成或推理，业务系统保留自身的交易与审批责任。

原则	实际含义	对业务的影响
数据库实体权威	身份、资源、任务、版本与回执有明确实体；JSON 承载有界内容与协议数据	可以查询真实状态，避免以聊天措辞判断完成
人承担责任	业务 Agent 绑定有效的人类主要责任人	可追溯负责人，并按当前组织和授权调整访问
权限在使用时生效	发现、读取、发送与披露分别核对当前权限	旧引用和缓存不能延续已撤销的访问
精确版本与来源	内容、计划和上下文绑定具体修订及摘要	可以区分历史依据与当前事实
自主执行有边界	执行依赖授权、租约、预算、停止条件与回执	恢复和重试不能自行扩大能力
未知状态如实保留	传输不明、缺失用量和不可用聚合有独立状态	不把未知当作成功或零成本

这些原则不能保证模型正确，但能帮助人检查输入、控制动作、定位错误和追溯责任。事实判断、业务风险和最终验收仍需要由相应人员负责。

3. 总体架构与权威数据模型

川序的应用层通过 FastAPI 服务提供 Web 与机器访问，React Dashboard 面向管理和工作操作，Portal 面向用户交互。业务服务复用同一套身份、资源、任务与授权逻辑；数据库适配层处理各底座的 SQL、索引、原生权限和检索差异。[R1]



图 1：川序的逻辑架构。数据库保存权威状态；模型、工具和外部 Agent 经当前授权及执行控制参与工作。图中层次用于说明职责，不代表每层必须部署为独立服务。

权威对象包括主体和组织关系、安全域与频道、Knowledge 与 Memory 修订、Task 与 Step、工作空间和分支、Graph 的计划与运行、Loop、Work 与交接，以及候选、审核、发布、租约和审计记录。对象之间的关联比一段自然语言摘要更适合表达状态变化、当前责任和访问边界。

属性图描述业务对象之间的关系，可用于有界关系检索和可视化。执行图则描述节点、依赖、屏障和运行结果。两者用途不同。属性图投影可以重建，既不覆盖原始实体，也不独立授予权限；执行图必须绑定已编译和发布的具体计划。

每个应用进程在导入数据库驱动之前固定后端，冲突配置在连接前拒绝。一个社区版包提供三种部署选择，每次安装仅使用所选底座及其依赖。跨底座切换需要独立目标、安装与迁移验证；这不是把三个数据库透明联合为一个在线事务系统。

4. 人、Agent、组织与安全域

Human、平台原生 Agent、外部 Agent 和数据库 Owner 承担不同职责。Agent 注册、激活、运行就绪和资源授权是不同状态。业务 Agent 必须关联有效的 Human 主要责任人；注册成功不会自动取得模型、工具、知识或执行权限。[R5]

组织表达管理归属。Human 有一个主要组织，也可以有临时的其他组织关系；组织树及其有效闭包参与当前访问计算。安全域表达协作授权边界，频道承载会话与工作沟通。加入频道只是加入对话，不能自动读取频道成员能访问的全部资料，也不能获得数据库行访问权。

一次知识辅助任务通常同时涉及发起 Human 和执行 Agent。可使用来源应满足二者当前权限的交集，并继续满足资源、安全域、生命周期和用途约束。执行器不能借助数据库 Owner、另一名用户或更换令牌绕过原请求的限制。

外部 Agent 通过受控注册与身份凭证接入，可使用一次性秘密或 Ed25519 证明，获得与实例绑定的短期令牌。执行记录关联凭证、实例、租约和 fencing 信息。管理员能够区分“描述了一个 Agent”“该 Agent 已验证身份”和“该实例现在可以执行这一动作”。

实际部署还应明确人员离职、组织变更、责任人停用和凭证泄露后的处理。撤销权限应影响后续读取与执行；历史记录保留当时发生的事实，但历史快照本身不能恢复当前访问权。

5. 数据库原生多信号检索

v5.0.3 在知识回答与任务上下文中组合五类逻辑信号：关系属性、向量相似度、全文、标签和有界图关系。关系属性处理产品、版本、责任或时间等精确条件；向量辅助语义召回；全文保留名称和术语；标签表达分类；图关系补充对象之间受限的联系。具体索引、查询方式及可用信号由底座与当前配置决定。[R2]

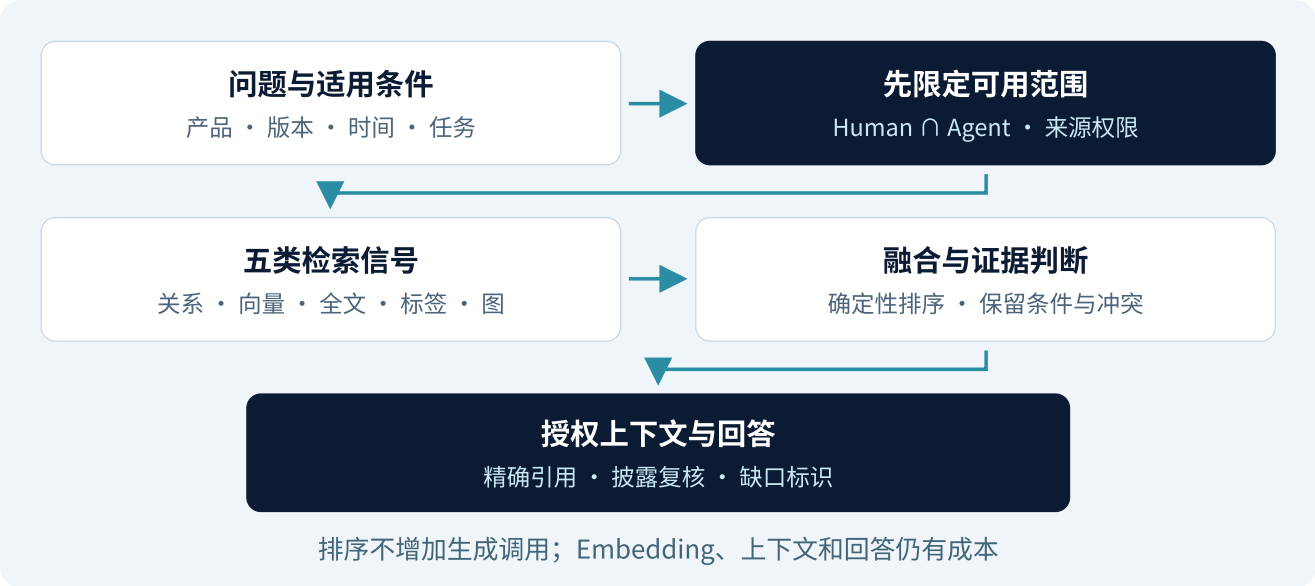


图 2：从问题到授权证据。权限与适用性先于候选截断；候选加载、模型披露及引用回读继续检查精确来源。

检索先限定主体、资源类型、生命周期、适用版本与时间，再在获准集合内召回、融合和截断。不能先取全库前若干条，再把无权记录删掉，否则既可能泄露信息，也可能挤掉真正可用的结果。排序得分不会抵消权限和适用性条件。

数据库查询与确定性排序融合不增加生成调用。同一契约中的查询向量可复用；重复内容可去重，但不同权限、条件、冲突和确切身份需要保留。查询 Embedding、上下文输入及最终回答仍消耗资源，不能把这一设计表述为“整个问答零 Token”。

检索服务或向量不可用时，仅允许使用已配置且经过授权的文本路径，并记录实际策略及原因。权限拒绝不能触发更宽的回退。系统还需分别判断完全匹配、部分匹配、无匹配和歧义；高分候选不直接等于充分证据。

答案可区分知识依据、知识摘录、模型补充、混合来源、需要澄清和知识不足。知识正文只发送给获准接收该资料的模型 Profile；未获准披露时，可以提供用户有权查看的摘录。需要通用模型补充时，必须由策略允许并单独标识。事实、推断和有条件建议应保持区别。[R3]

6. 有来源的知识入库与发布

川序的知识入库引导收集用途、来源位置、责任人、主题、适用产品与版本、时间、密级和目标受众，再提供分块预览及质量反馈。表头、代码、单位、否定和适用条件应随内容保留，避免分块后改变原意。[R4]



图 3：用户知识从来源到共享的生命周期。审核通过后的晋升产生私有原生实体；扩大共享范围仍需要独立、明确的授权。

预览不调用模型，也不保存或发布知识。缺少部分信息时可保留为草稿；缺项、重复和冲突反馈帮助补充材料，但不是事实真实性保证。候选进入已有的独立审核和发布链路，审核者不能自审自批，内容晋升和共享权限分别处理。

晋升得到私有原生实体，目标受众只是提交时的意图。组织共享或跨安全域传播需要额外授权。社区版保留通用候选与独立审核链；企业专属治理审批不应与这条公共链路混淆。

Web、API、MCP、CLI、策展导入和来源同步复用服务端规则。个人私有笔记保留已有直接写入路径；需要共享的直接创建会被引导到入库流程。受控来源同步产生候选，不因文件变更自动发布，更不会执行来源中的代码。

已发布知识应保留精确修订、摘要、原文位置、产品、版本和时间。更新、撤回和冲突处理都要作用到后续使用；摘要、回答历史和派生材料在读取或披露时仍需重新检查来源权限。内置产品学习资料另有受信来源绑定的发布机制，不能推广为任意用户内容均可自动发布。

官方事件与复盘草稿

v5.0.4 通过有界采集读取 GitHub、Cloudflare 官方状态接口，保存原始快照、事件时间、精确修订和幂等记录，沿用 Source、Task 与候选审核链。采集成功只表示外部资料已记录，不自动发布知识或执行修复。[R22]

故障复盘参考 PagerDuty 开放模板，区分事实、时间线、证据、缺口及待确认建议，并保留模板归属和许可。公开来源不能证明某家企业实际受影响，也不能填补未公布的根因。团队须补充内部日志、责任人和处置回执，再独立审核并显式晋升；示范报告保持草稿。

7. 时态记忆、工作便笺与上下文

Memory 适合保存需要跨任务使用的事件、观察和事实关系。事件发生时间、观察时间、入库时间和适用时间分别表达不同含义。迟到的数据不能因入库更晚而自动覆盖较早事实；没有时区或时间的信息保持未知，不补造看似精确的时间线。[R6]

冲突关系的确认与内容的发布是不同动作。恢复历史内容会创建新的后继版本，保留原历史，不把当前指针倒退为旧版本，也不借恢复降低密级。这有助于追溯一项判断何时被记录、何时被修订以及当时依据什么。

工作便笺保存已发现事实、放弃的方法、验证结果、未决问题和下一步。便笺及其来源有不可变修订，并受期限、Work 和 Handoff 权限检查约束。一次任务的中间想法不自动成为正式知识；复盘生成也先形成待审核候选，再由独立流程决定是否发布。

上下文资产采用先发现元数据、再加载精确版本的方式。绑定的来源和授权需要重新核对；预算必须预留系统提示、工具协议和输出空间。条件、否定、冲突及必要证据不能为了适配预算而静默截掉。预算不足时应阻止派发，缩小任务或调整已批准配置。[R7]

系统区分使用已知 tokenizer 得到的计数与保守估算，也区分“准备了上下文”和“有证据确认该输入已经发送”。这些差别使上下文诊断能够回答材料是否真正进入了那次模型调用，而不是只证明服务端曾经构建过它。

v5.0.4 的装配依据展示精确来源修订、表示方式、预算，以及选用、遗漏和降级原因。准备记录不授予来源权限，也不证明已派发。使用时继续复核来源、有效期和当前授权；来源撤销后，历史记录不能继续披露正文。该界面用于检查装配决策，不承担独立的传输字节证明。[R7, R22]

8. Task、Graph、Loop 与受控执行

Task 和 Step 提供稳定的工作身份，attempt 记录每次执行尝试。租约限定哪一个 Worker 当前有权工作，fencing 防止已过期执行器继续发布结果，幂等标识辅助识别重复请求。持久化状态使进程重启后的诊断和恢复有据可查。[R8]

Graph 将节点和依赖编译为具体版本的计划，经授权发布后，运行绑定该精确计划。节点可涉及 Agent、模型、Skill、工具、数据库、人、事件和子图；attempt、事件、检查点与结果分别记录。屏障保存参与者快照与到达记录，等待阶段不占用 Worker 执行租约。

Loop 在同一执行体系内提供有界迭代，具有次数、Token、时间和停止条件。评价可使用测试、差异、模型判断、人工、规格与聚合等方式；评价方式本身不改变结果的证据强度。Harness 将任务目标、环境和验收要求组织为运行蓝图，但蓝图不会授予模型或工具权限。

逻辑去重不能保证任意远程系统的副作用恰好发生一次。创建工单、发起支付或修改外部系统后若响应丢失，需要查验真实结果，必要时补偿或人工处理。离线重放是诊断工具，不能默认重做外部动作；迁移预检也不意味着支持任意活动运行的在线迁移。

处理预算不能通过重启进程、新建 attempt 或分叉 Graph 重置。需要恢复时，人应批准精确目标、版本和有界预算。接受恢复请求、真正派发、执行成功和业务批准分别记录；BLOCKED、DEGRADED 与 UNKNOWN 保持独立含义。

9. Work、Handoff 与 DB4A2A 协作

Work 记录工作目标、责任、版本与验收条件；Handoff 记录交接提议、接收确认、开始、结果及读取证据。默认单接收方在接受后承接责任。显式并行交接可限定为 2–16 个接收方，由协调者保留责任；这个边界是工作流约束，不是系统容量指标。[R9]



图 4：连续工作与责任交接。每一步都关联真实版本和回执；生成的总结不能代替接收确认或执行结果。

DB4A2A 是川序利用数据库共享上下文的协作模式。小型任务或状态信封传递 `task_id`、`context_ref`、`snapshot_digest`、`expected_version`、`scope_ref` 与 `branch_policy`，接收方再以自身身份读取获准的精确上下文。引用本身不授予访问权。[R10]

默认读取不改变来源。需要派生工作分支时，要明确授权，原始上下文、派发与审计保持一致；合并回原分支仍是独立动作。Task、Graph 和协作来源快照记录捕获时的状态，读取时重新核对来源权限，不能把旧快照当作当前运行状态。

DB4A2A 可以减少重复搬运同一份上下文，以及复制后出现的版本偏差。收益取决于数据共享条件、访问路径、上下文大小与模型使用方式，不能推导为统一的速度或 Token 节省比例。它也不是标准 A2A 的替代协议；没有共同的获准数据平面时，仍需要 A2A 的负载或工件传输方式。

对于外部发送，未知结果保持 `UNOBSERVED`，在确认原请求实际完成前不盲目重发。新签发令牌不能替代已经撤销的原凭证，为旧操作补造授权。这样，交接不仅传递“下一步做什么”，也保留“依据是什么、谁有权继续、前一步是否已完成”。

10. 工具、Skill、协议与框架接入

川序将能力发现和动作执行分开。Agent 可以先查看 Skill 或工具元数据，再按需要读取精确内容；执行仍受当前注册契约、凭证、安全域、来源权限与租约控制。Skill 文档中的指令和工具返回文本不能提升为系统安全指令。[R11]

接入类型	当前明确覆盖	使用边界
MCP	2025-06-18、2025-11-25 的版本化 Streamable HTTP 适配	按各版本的操作和契约使用，不宣称任意 MCP 服务兼容
后续 MCP 适配	2026-07-28 的有界 `server/discover`、`tools/list`、只读 `tools/call`	完整结果及受限 schema；不支持订阅、任务、MRTR、`input_required` 等未开放操作
A2A	1.0.1 的有界 HTTP JSON 任务流程及有限状态 SSE	不扩展为所有 SDK、gRPC 或推送传输的支持
框架运行	固定 LangGraph 1.2.12，锁定依赖摘要与经验证的 Linux rootfs	需要隔离运行证据；不等于任意版本与任意框架托管
可观测导出	OTLP/HTTP JSON 1.9.0 的已提交 trace 元数据	不包含提示词或回答正文，不等于所有 OTLP 传输实现

发现后的能力需要明确批准和暴露。重新发现可能改变契约，因此旧批准失效，后续执行需使用匹配的当前版本。具备某种适配器描述转换，也不能当作对应外部产品原生版本已经完成端到端认证。

发布包和 Skill 包可进行服务端签名验证，但必须存在独立信任的公钥；接收到文件和批准激活是不同状态。合作式执行包装仅能控制主动接入的进程，不能保证终止任意远端进程。客户接入自有工具前，还应确认幂等、超时、审计、回滚和真实副作用的查询方式。

11. 模型路由、容量、用量与费用

模型 Profile 保存具体配置与修订，并以实际探测记录区分文本、流式、结构化输出、工具、图像、推理参数、用量和取消能力。模型名称或类似 `/v1` 的地址不能证明这些能力。服务端校验的 `TEXT_JSON` 与精确业务 schema 探测过的 `NATIVE_STRICT` 也是不同契约。[R12]

受控只读工具回合是一个有意收窄的示例：使用一个当前注册的只读 GET 工具，具备人工注册与逐次动作授权，最多两次模型请求、一次工具调用、零次修复，并共享整体截止时间。GET 方法本身不能证明只读，仍要审核工具语义。此边界不能外推为任意自主工具循环都已安全开放。

模型网关与直接调用可以按 Profile 共存。共享推理资源需要同一个权威容量协调器，登记参与身份、地址别名和 Profile 版本；各数据库分别加锁不会形成跨库全局队列。源侧复核业务来源，协调侧复核中央容量和上游权限，双方不互相代授另一数据库的私有权限。[R13]

客户端断开只证明本地取消，不证明 Provider 已停止计算。结果未知的请求不能因租约过期就释放容量并重发，应先确认那个实际请求的完成情况。完整传输也不等于业务输出通过验收，拒答、截断和校验失败应分别记录。

用量台账保留来源、请求摘要、Provider 报告的输入、输出及可获得的缓存或推理维度，区分排队、首内容、请求与总时长。缺失用量标为不完整，不能计作零；未知价格或观测不全时费用保持未知。平台估算不是供应商账单，企业费用分摊需要明确范围及完整比例。[R14]

元数据台账不存提示词、回答和隐藏推理正文；另有显式配置的业务内容、评估或加密重放存储应单独管理。只读看板按权限展示 Agent、会话、Task、Loop 及有界周期的 Token 和费用概况；不可用聚合显示空值或降级，而非伪造零消耗。未经可信适配观测的外部直连流量仍可能不在统计范围内。

12. 安全、治理与合规职责

最终传输与输出边界

v5.0.4 的公共 native transport 检查最终 Provider JSON 的完整文本，并发送同一份序列化字节；完整输出继续执行校验。该路径帮助缩小“检查的是一份输入，发送的是另一份输入”的差距。托管图片像素语义和独立 DispatchTicket 字节审计仍有明确限制，因此不能据此声称全面 DLP 或所有适配器均获得同等覆盖。[R22]

川序组合应用层检查与数据库原生边界。会话、CSRF 防护、短期 Agent 凭证、资源授权、安全域和来源检查各有职责；Web 来源 IP 规则覆盖 IPv4 与 IPv6，环回紧急登录仍需认证。凭证保留在服务端加密配置或受限配置文件中，不能由客户端自行发放权限。[R5]

PostgreSQL 使用受限 Owner、专用登录角色、身份映射与强制 RLS，拒绝通过超级用户、BYPASSRLS 或 Owner 回退绕过。Oracle 依赖原生 End User、上下文、Data Role 与 Data Grant 等部署前提；这些前提也适用于社区版。YashanDB 采用独立用户、对象授权、受限投影与会话身份机制，不能照搬 Oracle 的安全语义。[R16]

企业版增加专属审批、审计管理视图、LDAP、治理与合规模块，可用于职责分离、策略、保留、法律保全及受限导出等场景。公共路径仍有必要的审计写入，不能将社区版描述为“没有审计”。通用知识候选的独立审核也不等于企业专属审批功能。OIDC、SCIM 等保留的类型名称不能当作已交付连接器。

运行隔离依赖实际主机条件：专用身份、namespace、cgroup v2、资源限制、seccomp、bubblewrap、LSM、只读根及网络策略等证据。缺少要求时应拒绝执行。当前已验证默认拒绝出网；选择性出网白名单没有相应实现和证据时继续拒绝。参考容器或虚拟机生命周期适配器返回未验证状态，不代表已自动创建并验证环境。[R17]

应用进程隔离不覆盖拥有宿主机 root 或基础设施管理权限的攻击者。数据库集群、操作系统、证书、密钥托管、备份、网络和安全事件响应仍需客户与实施方共同负责。合规 Agent 可以提出建议和证据，不能继承管理员权力；技术控制也不等于获得监管许可或独立认证。

13. 数据库适配与部署条件

三种底座遵循共同的业务对象和权限目标，通过适配器实现原生差异。选型应依据已有数据库能力、许可、DBA 经验、数据规模和治理要求；不能仅因逻辑接口相同，就假定三个数据库具有相同索引、执行计划或吞吐。[R1、R16]

底座	典型隔离与准备	部署时重点确认
Oracle	独立 PDB、应用 Owner、原生身份与授权准备	Text、向量等所需能力及版本；DBA 脚本；恢复后的对象、约束和授权
PostgreSQL	专用数据库、非超级用户 Owner、登录角色及强制 RLS	所需扩展与索引、会话身份、Owner 权限和越权拒绝
YashanDB	独立 PDB、用户与对象授权、受限数据投影	匹配的客户端与驱动、检索能力、会话身份和原生权限差异

正式包的运行基线是 x86_64 Linux、glibc 2.34 或以上，并以 RHEL / Oracle Linux 9.8 或以上及另行验证的兼容组合为依据。部署应使用与包内 wheel 和数据库驱动 ABI 匹配的 CPython 3.14 环境；不能把“更高 Python 版本”理解为二进制依赖自动兼容。[R18]

软件可携带离线依赖与数据库客户端，社区版只安装所选后端依赖。DBA 事先完成需要特权的准备，运行配置不保存 SYS 或 root 凭证。Oracle 和 YashanDB 所需的窄范围账户状态视图应由 DBA 明确创建，不通过宽泛系统目录授权替代。配置和密钥文件使用受限权限，替换安装时保留有效配置与密钥。

模型、Embedding 和外部工具可位于客户批准的服务端点。数据是否允许离开部署边界，取决于具体 Profile、来源披露规则、网络和合同安排。私有化安装本身不证明所有模型调用都在本地，必须检查实际路由与发送证据。

14. 安装、升级、运维与恢复

新装在明确的空目标中执行包内初始化 SQL，不依赖模型生成数据库结构，也不因安装成功自动启用敏感能力。上线前要确认实际产品、数据库、运行时与配置一致，检查数据库身份、迁移状态和权限负向用例。[R18]

升级应先备份并确认可恢复，再预检、执行支持的迁移链，核对已有数据、约束、权限和应用行为。历史已应用 SQL 与校验和保持不可变，修正通过新迁移完成；部分执行恢复必须依据精确迁移日志。结构对比有助于诊断，但不能代替真实升级，也不能只看终端显示迁移完成。

`/api/health` 说明进程存活，`/api/ready` 检查数据库与控制平面就绪，两者不应混用。日常观测应关联 Task、attempt、模型请求、工具回执、租约和迁移状态。执行未知、来源失效、预算不足与数据库不可用应有各自可解释的状态和处理路径。

数据库备份之外，还要保护加密配置、密钥、批准的 Skill 或运行镜像，以及迁移和签名验证记录。演练应检验恢复后的身份、知识精确修订、授权、任务状态和外部副作用，而非仅验证数据库能打开。恢复点目标与恢复时间目标由实际部署架构和演练结果决定。

应用多进程运行及租约恢复机制不等于数据库集群高可用。复制、故障转移、负载均衡、证书续期、容量规划和灾难恢复需在部署方案中明确。失去租约的旧执行器不能发布成功；外部操作未知时，应先查清结果再决定重试，避免恢复流程制造重复业务动作。

15. 版本、许可与学习资料

产品	数据库选择	公共能力	企业专属模块	许可
川序社区版	Oracle / PostgreSQL / YashanDB，单次部署选一	知识、记忆、任务、协作与公共执行控制	不随包交付	Apache-2.0
川序 Oracle 企业版	Oracle	公共能力及相应原生适配	有	BSL-1.1，生产需商业许可
川序 PostgreSQL 企业版	PostgreSQL	公共能力及相应原生适配	有	BSL-1.1，生产需商业许可
川序 YashanDB 企业版	YashanDB	公共能力及相应原生适配	有	BSL-1.1，生产需商业许可

企业专属模块包括 approvals、audit、ldap、skill_token、orchestrator、governance 和 compliance，社区包在构建时排除这些模块。表中公共执行控制包括公共 Task、Graph、Loop 等能力，不能仅凭“编排”一词推定全部图执行只属于企业版。数据库原生身份与权限前提不因采用社区版而省略。[R15]

v5.0.3 提供九个完整的中英文平台学习单元。社区版保留学习入口与四类简要产品说明，软件包不携带也不自动安装九个完整单元。有权管理员可从包外独立 JSON 文件显式导入与产品、版本和底座匹配的资料。企业版随包携带完整资料并自动部署。升级保留已安装知识、旧快照和用户修改，不静默删除。[R19]

学习内容采用角色与目标目录，Agent 先发现元数据，再按摘要读取精确正文。资料或 Skill 说明不会授予执行权限；JSON 摘要一致性也不等于独立的发布者签名认证。

社区版遵循 Apache-2.0，并需保留相应通知与第三方许可。企业版按交付的 BSL-1.1 文件及商业合同使用；当前许可证列明转换日期为 2028 年 5 月 28 日、转换许可为 Apache-2.0。数据库、模型、客户端和其他依赖的许可应分别核查。本文不增加产品支持义务，也不把未来可选的交付形态表述为现有能力。

16. 应用场景与工作方式

以下为说明性场景，均应在客户自己的资料、模型和授权下验证。

技术支持与运维知识。 Human 提交包含产品和版本的问题，Agent 检索适用手册、已批准变更与故障记录。回答保留名称、版本、原文位置和前提；证据冲突时提出澄清。若需诊断工具，先执行受控只读动作；修改生产系统另走批准与外部执行流程。价值评估可观察找资料时间、来源覆盖和人工复核负担。

研发问题的持续交接。 第一名 Agent 建立 Task，记录复现、排除方法和剩余问题；Handoff 将精确来源及已批准决策交给另一名 Agent。接收者以自身身份读取，确认接受后继续工作。Work 验收依据实际测试和操作回执。完成后，适合复用的结论进入知识候选，尚未确认的推断保留在便笺。

跨团队政策与流程问答。 不同组织可共享同一服务入口，但只在各自权限范围读取资料。旧版制度、适用日期和例外条件必须保留；频道成员关系不扩大制度访问权。涉及合规解释或权利义务的结论交由指定人员确认，平台提供可核对的依据和操作记录。

受控自动化与人机协作。 将既有工作流表示为 Task 或版本化 Graph，对每类工具、模型、数据库和人工节点规定输入、批准、预算及结果标准。非幂等动作产生未知结果时先查询外部状态，再由授权者决定恢复或补偿。自动化范围从可观察、可撤销的动作逐步扩展。

这些场景的共同价值假设是减少重复传递信息和查验状态的工作，并增强责任、权限和来源的可核查性。能否转化为成本或效率改善，需要以真实业务基线和试点结果判断。

17. 验证方法与当前证据边界

川序将平台数据正确性与模型语义质量分层检查。权限、来源版本、时间、任务状态、实际输入、引用和发布属于可逐项验证的控制；模型在正确输入上仍可能漏掉条件或作出错误推断。传输成功不能代替语义验收。[R20]

v5.0.4 的验证覆盖事件来源、复盘候选、上下文装配依据、最终传输与输出检查，以及五个交付物、七个数据库目标的十四条新装和真实升级路径。既有检索、入库、学习和权限链路按依赖核对保留有效结果，受影响部分单独补测；不能称所有最终文件都重新执行过全部测试。

本版对显式来源与混合检索的同轮准备路径进行了成本测量，保留使用时的授权重验。局部耗时不能外推为 Token、费用或普遍性能收益，也没有启用新的 query/TTL 缓存。共享编码仅完成实验设计，默认关闭。模型摘要仍要逐项核对来源，结构有效不能代替事实、根因或业务验收。[R22]

检索验证使用预先冻结的问题与不同规模的干扰集合，核对所需原文是否召回、条件是否进入有界上下文、引用是否精确以及无权内容是否泄露。这些结果建立了选定语料和配置中的行为证据，不代表客户全量资料覆盖或任意规模容量。关键词对照的漏召回以及模型回答中的语义失败和遗漏继续保留。

已有证据类别	可以说明	不能据此推定
原生权限及业务链路检查	所选身份与流程下的允许、拒绝和状态变化	独立安全审计、全威胁覆盖或监管认证
冻结语料检索与上下文验证	所需来源、条件、引用和授权在相应输入上的表现	所有问题正确、所有底座同等性能
新装与稳定状态副本升级	选定来源、迁移链和对象权限的实际行为	任意历史客户库可直接升级、完整物理 PDB 等价
自动浏览器与凭证脚本	界面路径、身份和操作回执可执行	真人新手研究或自主 Agent 端到端验证
模型调用和输出检查	所用配置下的真实输入、传输与具体输出	生产模型已全面评估或模型语义全部通过

当前发布结论是在批准的范围内成立，并保留模型错误及未评估项。客户仍需验证自己的数据库版本、权限、数据、模型、并发和运维条件。本文不提供统一延迟、吞吐、Token 节省比例或 SLA；内部验证数字也不应直接成为采购性能承诺。

18. 概念验证、选型与实施建议

概念验证应从一个有明确责任人、可取得资料并可核验结果的场景开始。先约定验收口径，再配置产品和环境，避免把“能够对话”误认为“已经可用于业务”。下表是建议的分阶段方法，不限定实施时长。[R21]

阶段	主要工作	建议退出条件
场景与数据准备	明确目标、责任人、数据分类、授权矩阵、基线与问题集	资料与许可可用，预期结果和拒绝用例获确认
环境与控制验证	部署所选底座，验证身份、RLS 或原生权限、模型路由与撤权	无权读取被拒，模型披露和任务证据可追溯
业务闭环验证	执行入库、问答、Task、交接和人工批准；测试未知与恢复	不仅得到回答，还能核对来源、责任、结果与异常处理
容量与上线评审	测试实际规模及并发，演练升级、备份、故障和恢复	质量、容量、运维和许可达到事先约定标准

建议同时测量检索召回、精确引用、条件保留、权限拒绝、语义正确率和人工复核量；性能记录区分数据库查询、Embedding、排队、首内容、完整回答与外部工具时长。使用固定语料、问题、版本和模型配置，报告分位数、失败与超时，不只展示最好的一次。

总成本应包括数据库与软件许可、主机与存储、模型及 Embedding、运维、知识治理和人工审核。排序不增加生成调用可以作为设计事实，但投入产出仍须用实际工作量、质量与风险变化计算。不要把文本缩短等同于同等质量的成本节省。

企业负责业务目标、数据合法性、人员与资源授权、风险接受和上线决定；实施方负责约定范围内的适配、部署、迁移和运维验证；DBA 与安全团队负责原生边界及基础设施；业务审核者确认知识和结果。选型时应确认企业治理模块需求、后端前提、升级来源、支持范围及合同，而不是以一次演示替代生产评审。

附录 A：术语表

术语	本文含义
Human / Principal	人用户；主体是可认证和授权对象的共同身份概念
Agent	与责任人、实例和当前授权关联的执行主体
Security Domain / Channel	协作授权边界 / 会话与沟通空间；二者不互相替代
Knowledge / Memory	经治理复用的知识 / 保留来源和时态的记忆实体
Task / Step / attempt	工作、步骤与具体执行尝试
Work / Handoff	工作责任与验收对象 / 有回执的责任交接
Graph / Loop / Harness	版本化执行图 / 有界迭代 / 运行与验收蓝图
属性图投影	从权威实体构建的关系视图，不授予资源权限
Profile	有明确修订、配置和实际能力证据的模型接入配置
Lease / fencing	执行权的时间限制 / 拒绝过期执行器写入的机制
DB4A2A	利用数据库引用传递获准精确上下文的协作模式
MCP / A2A	工具和上下文接入协议 / Agent 间任务协作协议；支持范围依版本而定
摘要 / 签名	内容一致性标识 / 需信任密钥验证的来源证明；两者作用不同
RLS / PDB	行级安全 / 可插拔数据库，按所选底座使用
UNKNOWN / UNOBSERVED	结果未知 / 缺少可信观测，均不能当作成功

附录 B：资料依据与继续阅读

以下资料按主题列出，以 v5.0.4 所选产品交付包内适用文件为准。中文指南通常以 `_zh.md` 结尾，对应英文指南去掉 `_zh`。文件名用于在交付包中定位，不要求离线读者访问研发系统。

编号	资料	主要用途
R1	`architecture_zh.md`	总体架构、数据与权限权威
R2	`channel-knowledge_zh.md`	知识检索、上下文和来源检查
R3	`answer-quality_zh.md`	证据充分性、来源类别及回答边界
R4	`knowledge-intake_zh.md`、`curated-knowledge_zh.md`	入库引导、候选、独立审核与发布
R5	`security_zh.md`	人与 Agent 身份、会话、安全域和访问边界
R6	`memory-working-notes_zh.md`	时态记忆、便笺、修订和复盘
R7	`agent-context-assets_zh.md`、`workspace_zh.md`	精确上下文、预算与分支
R8	`graph-engineering_zh.md`、`loop-engineering_zh.md`、`harness_zh.md`、`processing-recovery_zh.md`	图、循环、运行契约与恢复
R9	`continuity-operations_zh.md`	Work、交接和当前来源授权
R10	`db4a2a_zh.md`	数据库上下文引用协作
R11	`agent-extensions_zh.md`、`mcp-versioned-interoperability_zh.md`、`framework-agent-gateway_zh.md`、`release-signing_zh.md`	扩展、协议、框架和签名边界
R12	`model-workflows_zh.md`	Profile、结构化输出与受控工具回合
R13	`model-endpoint-capacity_zh.md`	共享容量、请求生命周期与未知结果
R14	`model-usage-and-wallboard_zh.md`	用量、费用与可观测范围
R15	所选包的 `LICENSE`、版本类别与许可说明	社区版和企业版的功能与使用范围
R16	`minimum-privileges_zh.md`	数据库原生权限和 DBA 准备
R17	`runtime-isolation_zh.md`	主机门槛、隔离证据与拒绝行为

编号	资料	主要用途
R18	`deployment_zh.md`、`python-runtime_zh.md`、`linux-platform-compatibility_zh.md`	新装、升级、依赖与平台前提
R19	`platform-learning-delivery_zh.md`	学习入口、独立 JSON 与升级保留
R20	v5.0.4 版本说明及交付范围说明	当前功能、验证口径和保留边界
R21	`poc-readiness_zh.md`	场景、部署与验收准备
R22	本版事件接入与上下文说明；独立研究资料 `docs/v5.0.4/enterprise-incidents/` 中的模板、ATTRIBUTION 与 PagerDuty 许可	官方来源、报告草稿、装配依据、测量与最终传输边界；独立研究资料不视为软件包内置内容

许可证、第三方通知和商业合同具有各自适用范围。Oracle、PostgreSQL、YashanDB、LangGraph 等名称属于相应权利人；本文提及兼容范围不表示这些机构为川序背书或认证。