

TDSQL Nexa

产品简介



腾讯云

【 版权声明 】

©2013–2026 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

【 商标声明 】



及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

【 服务声明 】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

【 联系我们 】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或 95716。

文档目录

产品简介

产品概述

产品优势

应用场景

应用场景概述

智能驾驶训练

产品简介

产品概述

最近更新时间：2026-10-16 11:30:00

TDSQL Nexa 是一个 AI 原生多模数据库：它在一套系统内融合了宽表、分析、搜索等计算能力，并通过核心引擎与实时数据流通道，把各引擎的数据无缝打通、把各引擎的能力按需组合编排，从而支撑智能驾驶、具身智能、搜索、广告、推荐以及企业 Agent 等场景下，从数据产生、分析到应用的全链路。

为什么选择 TDSQL Nexa

长期以来，构建一个数据应用意味着先做一道选择题：交易数据放进事务数据库，日志与文本丢给 Elasticsearch，报表分析再搭一套数仓，语义检索还要额外引入向量库。每一种负载都能找到一个最合适的数据库，但把它们拼在一起，得到的却是一套离散、碎片化的数据方案——数据被复制成多份，散落在不同系统中，靠 ETL 和消息队列艰难地缝合。系统复杂度、运维成本与故障概率，都随着组件数量的增加而被成倍放大。

进入 Agent 时代，这种碎片化的代价被进一步放大。Agent 需要的是统一、一致的上下文，以及一套可以直接调用的数据处理工具；而面对四五套语义各异、口径不一、时效参差的数据系统，Agent 既难以获得完整的数据视图，也难以可靠地完成数据操作。数据层的割裂，最终会拖慢整个业务的迭代与演进速度。

在 TDSQL Nexa 里面，事务、分析、检索等多种存储与计算引擎开箱即用，覆盖数据应用所需的主要负载类型；更重要的是，这些引擎在产品内原生打通，数据在引擎之间实时流转，能力之间可以自由组合编排。业务不再需要为了一条数据链路去拼装多套系统，而是可以在一个数据库里完成整个数据应用的构建。

在此之上，TDSQL Nexa 还内置了 Knowledge Agent，帮助人与 Agent 更好地理解数据——数据在哪里、是什么含义、该怎么用，从而让数据应用的构建过程本身也变得更加智能和高效。

TDSQL Nexa 的架构



TDSQL Nexa 的架构中最核心的是多模态计算引擎。由此向下看的组件解决数据如何流动与沉淀，向上看的组件解决能力如何被编排和使用。

核心：多模态计算引擎

每一类计算引擎都依托各自的底层存储结构，提供一类专门的数据处理能力，同时保持对主流开源生态的接口兼容。

引擎	核心能力	接口兼容
宽表引擎	高并发写入与点查、ACID 事务、宽表存储	MySQL、HBase
分析引擎	分布式并行计算分析、高效数据压缩与扫描	MySQL、PostgreSQL
搜索引擎	大规模向量与文档数据的存储与检索	OpenSearch、Elasticsearch
AI 计算引擎	非结构化数据处理：文档解析拆分、视频拆条与转码等	Ray

向下：NTS 引擎与统一存储

有了核心的多模引擎之后，往下一层是 NTS 引擎，它是各引擎之间数据得以无缝打通的关键：

- 异构索引同步：负责各计算引擎之间异构索引的构建与同步，让同一份数据在不同引擎中保持一致视图。
- 实时入湖：将各多模存储引擎的数据实时沉淀到数据湖，形成低成本、格式开放的统一底座（基于 COS，采用 Lance、Iceberg 等表格式）。

- 离散数据纳管：对外部离散的各类数据源提供纳管能力，一键同步到各异构引擎，或直接写入数据湖。

湖上数据的访问性能，由统一分布式数据加速层 GooseFS / CFS 提供，弥合低成本对象存储与高性能计算之间的性能差距。

向上：核心引擎&统一应用接入接口

在多模引擎之上是核心引擎。它让多个引擎对外表现为一个数据库或者是一个表：各引擎既可以作为彼此的索引相互配合，也可以被组合编排后统一对外提供服务——应用只需表达需求，无需关心由哪个引擎执行、如何协同。再往上是兼容各开源协议的访问层，提供 SQL、HTTP 与 MCP 等接入方式，既贴合既有的开发习惯与工具生态，也为 AI Agent 提供标准化的数据调用入口。

最上层：企业级数据智能服务

构建在整套数据系统之上的企业级数据智能服务，让数据不只是被存储和查询，更能被理解和使用：

服务	能力
Knowledge	基于统一的数据解决数据洞察、智能决策问题，沉淀专家经验
Studio	统一的 SQL 编写、作业开发与作业调试能力
Trace	诊断分析、链路追踪、日志管理
Catalog	血缘跟踪、权限管理、资产盘点

产品优势

最近更新时间：2026-10-16 11:30:00

TDSQL Nexa 将宽表、分析、搜索等多模计算能力融合在一套数据库系统中，各引擎原生打通、能力按需组合编排，帮助您以更简单的架构和更低综合成本构建数据应用。本文介绍 TDSQL Nexa 针对的核心痛点、对应的解法，以及各引擎的特色能力与性能表现。

行业痛点：碎片化的数据架构

长期以来，构建一个数据应用意味着先做一道选择题：交易数据放进事务数据库，日志与文本交给 Elasticsearch，报表分析再搭一套数仓，语义检索还要额外引入向量库。每一种负载都能找到一个合适的数据库，但把它们拼在一起，得到的却是一套离散、碎片化的数据方案：数据被复制成多份，散落在不同系统中，靠 ETL 和消息队列艰难地缝合。系统复杂度、运维成本与故障概率，都随着组件数量的增加而被成倍放大。进入 Agent 时代，这种碎片化的代价被进一步放大。Agent 需要的是统一、一致的上下文，以及一套可以直接调用的数据处理工具；而面对多套语义各异、口径不一、时效参差的数据系统，Agent 既难以获得完整的数据视图，也难以可靠地完成数据操作。数据层的割裂，最终会拖慢整个业务的迭代与演进速度。

TDSQL Nexa 的解法：一个数据库承载完整数据应用

- **多模统一引擎**：宽表、分析、搜索与 AI 计算等引擎开箱即用，覆盖数据应用所需的主要负载类型。业务无需为一条数据链路拼装多套系统，可以在一个数据库里完成整个数据应用的构建。
- **单事务多模写入**：多模数据可在单个事务内完成写入，从源头保障各引擎间的数据一致性，免除跨系统同步链路及其带来的一致性风险。
- **原生打通，免除 ETL**：各引擎在产品内原生打通，数据变更经 NTS 引擎实时流转，写入即可被检索、可被分析；各引擎既可以作为彼此的索引相互配合，也可以组合编排后统一对外提供服务，应用只需表达需求，无需关心由哪个引擎执行、如何协同。

各引擎特色能力

搜索引擎

TDSQL Nexa 提供高性能、高可靠的分布式搜索引擎，兼容 OpenSearch、Elasticsearch 标准接口，支持 PB 级向量与文档数据的存储与检索。

- **分布式架构，水平扩展**：数据按分片水平拆分、多副本部署，支持在线扩缩容，存储容量与读写吞吐随节点增加线性提升。
- **多类索引协同**：提供面向高维向量的近邻索引（支持 HNSW、IVF 等索引算法与多种量化压缩方式），以及倒排索引、列存正排与多维数值索引，支撑语义检索、全文检索、范围查询与多条件组合过滤。
- **与多模引擎无缝打通**：与分析引擎组合，可在检索结果之上直接完成复杂聚合与多维分析；与宽表引擎组合，可为高并发在线业务补齐事务处理与点查能力；数据变更实时同步，写入即可被检索。

- **生态兼容**：提供兼容 Elasticsearch、OpenSearch 的 HTTP 接口与 SQL 接口，既有应用与工具链可平滑迁移。

AI 计算引擎

AI 计算引擎是全托管数据处理引擎，向上提供作业管理与 workflow 编排、分布式计算运行时，向下统一纳管 CPU/GPU 异构算力。

- **全托管，开箱即用**：集群的搭建、扩缩容、升级与故障恢复均由平台托管，提交作业即可获得弹性算力，无需自行运维计算集群。
- **与多模引擎原生打通**：直接读取库内多模存储中的原始数据，处理产生的向量、标签与元数据可直接写入搜索引擎与宽表引擎，数据加工与数据服务在同一套系统内衔接。
- **多模态数据处理算子**：面向非结构化数据提供开箱即用的分布式处理能力，覆盖文档解析拆分、视频抽帧与拆条、音频转码、数据去重、质量过滤、敏感信息脱敏、模型推理打标签与 Embedding 化等常见数据加工环节。
- **全链路任务编排**：以 Kubernetes 原生方式串联数据处理、训练、评测与发布环节，形成可复用、可版本化的流水线，并配套全链路的运行监控与诊断分析。

宽表引擎

宽表引擎面向海量结构化与半结构化数据，单表可支撑万列、千亿行规模，在百万级并发下保持毫秒级的数据变更与查询性能。

- **多主写入，弹性扩展**：每个节点均可承担写入，集群吞吐随节点数线性提升；扩缩容透明在线，规模可从双节点平滑扩展至数百节点。
- **动态列，灵活承载稀疏数据**：单表支持万列级动态列，新增列无需预定义表结构或执行 DDL，未赋值的列不占用存储，适合标签、特征、元数据等宽而稀疏的数据。
- **亲和性调度，降低事务开销**：将主键、索引与关联表数据智能编排至同一节点，使原本需跨节点协调的分布式事务退化为单机事务，兼顾一致性与性能。
- **资源对等，成本更优**：节点角色对等，避免备节点资源闲置；存储引擎结合智能压缩，以更高压缩率承载海量数据。

分析引擎

分析引擎面向大数据量复杂查询，采用列式存储与高压缩比，提供兼容 MySQL 的访问接口，可直接承载复杂报表、多维聚合等分析场景。

- **分布式向量化并行计算**：查询被拆分为并行子任务在多节点间同时执行；节点内采用向量化执行，按列批量处理数据并充分利用 CPU 的 SIMD 指令与缓存局部性，加速大数据量下的复杂查询。
- **完全存算分离，多级缓存加速**：计算与存储彻底解耦，计算资源可独立弹性伸缩而无需搬迁数据；构建“对象存储—本地磁盘缓存—内存”多级缓存结构，热数据逐级上浮，兼顾对象存储的低成本与接近本地存储的访问性能。
- **生态开放**：底层存储采用 Iceberg 开放表格式，数据以开放标准落盘，第三方引擎可直接访问同一份数据，避免数据孤岛与格式锁定，让数据自然融入既有数据生态。

性能表现

搜索引擎

以下为基于 VectorDBBench 的测试结果，数据集 openai_large_5m（500万条、1536维向量），测试机型 64核/128GB内存。

产品与配置	向量索引大小	Top-100时延	Top-100召回率	Top-1000时延	Top-1000召回率
Nexa HNSW	30G	5.4ms	95.70%	22.6ms	95.30%
Nexa HNSW + SQ1	0.72G	5.4ms	91.60%	28.0ms	94.70%
Nexa HNSW + SQ16	16G	6ms	95.30%	25.2ms	95.30%
PostgreSQL HNSW	38G	12.3ms	96.40%	41.6ms	95.20%
Milvus HNSW	35.4G	8.7ms	99.60%	21.6ms	99.90%

Nexa HNSW 的 Top - 100 平均时延为5.4ms，低于同测试中的 Milvus（8.7ms）与 PostgreSQL（12.3ms）；借助 SQ1 量化，向量索引可从30G压缩至0.72G（约42倍压缩比），Top - 100检索时延保持不变（5.4ms）；SQ16量化下召回率与时延几乎无损。业务可按存储成本与召回精度的权衡灵活选择索引配置。

AI 计算引擎

指标	表现	实现方式
作业启动延时	P90小于10秒	底层资源提前预热，高频镜像自动缓存
作业提交吞吐	100 Job/s	提交链路解耦，多消费者并行拉取
单地域作业规模	10万作业稳定运行	持久化状态管理，分层队列降低大规模任务的管理开销

作业从提交到开始执行的 P90 延时控制在10秒以内，减少调试迭代中的等待；作业提交吞吐可达100 Job/s，可承载大批量数据处理任务的集中提交；单地域可支撑10万级作业稳定运行，任务规模增长时管理开销保持可控。

宽表引擎

TPC-C 测试（16核/32GB内存机型）：

线程数	tpmTOTAL	tpmC（NewOrders）
-----	----------	-----------------

64	208,948.14	94,035.97
1,024	434,737.57	195,442.64

在64 – 1,024线程测试范围内呈上升趋势：64线程下 tpmC 达9.4万，1,024线程下提升至19.5万，高并发下性能表现稳定无回落。

YCSB 测试（4节点、每节点16核/32GB，3亿行 × 100列）：纯读场景（Workload C）128并发下 QPS 约 6.8万、平均延迟约1.9毫秒；单列写入场景1,024并发下 QPS 约10万、平均延迟约10毫秒，海量数据规模下保持稳定的读写表现。

分析引擎

TPC-H 测试（3节点、每节点32核/64GB）：22条查询总耗时24.732秒，平均单条耗时约1.12秒，在复杂多表关联与聚合场景下保持稳定的低延迟表现。

说明：

以上性能数据均在特定测试环境与数据规模下测得，实际表现因业务负载、资源配置与环境差异而不同。测试时间：2026年9月。

痛点与收益对照

行业痛点	TDSQL Nexa 的解法	收益
多套数据库堆叠，架构复杂	多模统一引擎	在泛训练、企业级 Agent、搜广推等场景下，架构简化，综合拥有成本下降
数据孤岛与一致性问题	单事务内多模写入，引擎原生打通	免除跨系统同步链路，多引擎保持一致数据视图
AI 检索效果与工程落地割裂	向量与全文混合检索，内置多模态数据处理算子	检索召回与精度可灵活权衡，数据加工与数据服务库内闭环
规模增长后的性能与成本压力	分布式弹性扩展，结合存储压缩优化	容量与吞吐随节点线性扩展，量化压缩与列存高压缩比降低存储成本
企业级安全与合规要求	企业级数据安全和治理体系	血缘跟踪、权限管理与资产盘点统一管控

应用场景

应用场景概述

最近更新时间：2026-10-16 11:30:00

TDSQL Nexa 在一套系统内融合宽表、分析、搜索与多模数据处理能力，支撑从数据产生、分析到应用的全链路。本文介绍 TDSQL Nexa 的典型应用场景、各场景的数据形态与关键能力，并提供场景选型速查表，帮助您结合业务需求选择合适的使用方式。

场景一：企业知识库与 RAG 问答

业务描述：企业通常需把规章制度、产品资料、客服工单、研报等分散的文档沉淀为统一知识库，并通过检索增强生成（RAG）构建智能问答、智能客服与智能投研等应用。传统方案需将向量库、关系数据库与全文检索系统拼装在一起，数据在多套系统间靠 ETL 同步，链路长、口径易不一致。

数据形态：原始文档、切片、向量、元数据。

关键能力：

- 混合检索：向量近邻检索、倒排全文检索与标量过滤在同一搜索引擎内完成，支持 HNSW、IVF 等索引算法与 BQ、PQ、SQ 等量化压缩方式。
- 过滤条件下推：过滤条件在检索过程中下推执行，保障大规模数据下的检索效率。
- 增量更新：文档经 AI 计算引擎完成解析拆分与 Embedding 化后写入，数据变更经 NTS 引擎实时同步，写入即可被检索。
- 接口兼容：提供兼容 Elasticsearch、OpenSearch 的 HTTP 接口与 SQL 接口，既有应用与工具链可平滑迁移。

收益：文档、切片、向量与元数据在同一系统内闭环管理，免除跨系统同步链路；知识检索与知识库的数据管理能力同源，数据视图一致；兼容开源接口，迁移成本低。

场景二：智能推荐与个性化

业务描述：零售电商、内容平台等业务需要在海量商品与内容中实现个性化推荐：一方面要存储并频繁更新用户画像、物品特征等宽而稀疏的标签数据，另一方面要基于向量相似度完成候选召回，并对用户实时行为快速响应。

数据形态：用户与物品特征（标签、画像）、行为事件、向量。

关键能力：

- 宽表引擎单表支持万列级动态列，新增标签只需写入新列，无需预定义表结构或执行 DDL。
- 多主架构支持高并发写入，集群吞吐随节点数线性提升，从容应对大促等流量峰值。
- 搜索引擎提供高性能向量检索能力，支撑相似内容召回。
- 数据变更经 NTS 引擎实时同步，用户新产生的行为数据写入后即可参与召回。

收益：特征存储与召回检索在同一系统内闭环，免除特征在多套系统间的同步链路；标签体系可随业务灵活扩展；实时行为即时生效，支撑个性化的推荐体验。

场景三：语义搜索与内容检索

业务描述：站内搜索、内容检索等业务既要支持精确的关键词匹配，也要理解用户的语义意图，使同一含义的不同表述也能命中目标内容，并支持在检索结果上叠加条件筛选与分析。

数据形态：文本与内容数据、向量、结构化属性。

关键能力：

- 向量索引：支持 HNSW、IVF 等多种索引算法与多种量化压缩方式、距离度量，可在召回精度与存储成本之间灵活权衡。
- 倒排索引：面向文本的分词倒排结构，支持多语言分词与自定义词典，提供多种查询匹配模式与排序规则。
- 多维数值索引：支撑范围查询与多条件组合过滤，满足"语义相似加条件筛选"的组合诉求。
- 检索结果可直接交由分析引擎完成复杂聚合与多维分析。

收益：关键词检索与语义检索在同一引擎内完成，无需分别维护全文检索与向量检索两套系统；具备 PB 级向量与文档数据的存储与检索能力；兼容 Elasticsearch、OpenSearch 接口，既有搜索应用可平滑迁移。

场景四：AI Agent 记忆与上下文存储

业务描述：Agent 应用需要统一、一致的上下文，以及一套可以直接调用的数据工具。对话记忆、长期记忆、工具调用产生的结构化记录、Agent 运行轨迹等多类数据，若分散在多套语义各异、口径不一的数据系统中，Agent 既难以获得完整的数据视图，也难以可靠地完成数据操作。

数据形态：会话与记忆数据（向量与关系）、Agent 运行轨迹、工具调用元数据。

关键能力：

- 搜索引擎既可作为 Agent 应用的检索底座，也可承载 Agent Trace 等运行时数据。
- 宽表引擎提供高并发点查与 ACID 事务，支撑记忆与业务数据的可靠读写。
- 访问层提供 SQL、HTTP 与 MCP 等接入方式，为 AI Agent 提供标准化的数据调用入口。
- 核心引擎让多个引擎对外表现为一个数据库，应用只需表达需求，无需关心由哪个引擎执行、如何协同。

收益：Agent 通过统一接口获得一致的数据视图与数据操作能力；记忆检索与业务数据处理在同一系统内完成；运行时可观测数据与应用数据同库存储，便于回溯与诊断。

场景五：多模态内容管理（图文音视频）

业务描述：智能驾驶、具身智能等模型训练场景，以及传媒内容行业的素材管理，都需要对图片、文本、音频、视频等非结构化数据进行规模化加工：解析、切分、去重、脱敏、打标签、向量化，再将产物用于检索与分析。

数据形态：多媒体原始文件、解析产物（切片、帧、转码结果）、标签、向量。

关键能力：

- AI 计算引擎提供开箱即用的分布式处理算子，覆盖文档解析拆分、视频抽帧与拆条、音频转码、数据去重、质量过滤、敏感信息脱敏、模型推理打标签、Embedding 化等常见加工环节。
- 作业运行于全托管的弹性算力之上，统一纳管 CPU 与 GPU 异构资源，无需自行运维集群。
- 处理产生的向量、标签与元数据直接写入搜索引擎与宽表引擎，写入即可被检索。
- 处理结果可经实时入湖沉淀为 Lance、Iceberg 等开放表格式，供其他引擎直接读取。

收益：数据加工与数据服务在同一系统内衔接，省去外部 ETL 链路；处理产物无需在系统间搬运，写入即可被检索；湖上数据采用开放格式，避免数据孤岛与格式锁定。

场景六：风控与实时分析

业务描述：金融风控、实时运营分析等业务既要保障在线交易与事件的高并发可靠写入，又要对新流入的数据即时开展复杂的多维分析，同时还需要基于实体画像与相似度计算识别风险。

数据形态：交易与事件流水（结构化数据）、实体画像与标签、向量。

关键能力：

- 宽表引擎提供 ACID 事务与高并发写入点查，保障在线业务的可靠性。
- 分析引擎采用列式存储与分布式向量化并行计算，支撑复杂报表与多维聚合分析。
- 多维数值索引支撑范围查询与多条件组合过滤，配合向量检索完成实体相似度计算。
- 数据变更经 NTS 引擎实时同步，新数据写入后即可参与分析。

收益：在线事务与分析能力在同一系统内闭环，免除传统的跨系统 ETL 链路；风控所需的结构化过滤与语义相似计算可组合完成；分析引擎兼容 MySQL 接口，既有 BI 与报表工具可直接接入。

场景选型速查

场景	主要数据模型	关键能力
企业知识库与 RAG 问答	向量 + 关系 + 全文	混合检索、过滤条件下推、增量更新
智能推荐与个性化	宽表（标签与特征）+ 向量	动态列、多主高并发写入、实时同步与召回
语义搜索与内容检索	向量 + 全文	向量索引与量化压缩、倒排全文、多维条件过滤
AI Agent 记忆与上下文存储	向量 + 关系 + 运行日志	混合检索、MCP 接入、高并发点查
多模态内容管理	多媒体文件 + 标签 + 向量	多模态处理算子、全托管弹性算力、实时入湖
风控与实时分析	关系 + 宽表 + 分析列存	ACID 事务、分布式并行分析、实时同步

相关文档

- [快速入门](#)
- [产品概述](#)

智能驾驶训练

最近更新时间：2026-10-16 11:30:00

一、背景

智能驾驶正由功能验证走向规模化应用，数据已成为算法迭代、车辆运营和业务创新的核心生产要素。车端持续产生图像、点云、CAN 信号、事件日志和座舱交互等多模态数据；云端则需支撑样本检索、标注训练、仿真回放与运营分析。

面对车队规模增长和数据类型持续扩展，传统由多套数据库、搜索系统与数仓拼接而成的架构，难以兼顾处理效率、数据治理与长期运维成本。TDSQL Nexa 以一体化多模数据能力，为智能驾驶训练与分析链路提供统一的数据底座。

二、智能驾驶领域的管理痛点

1. 高并发写入、动态列扩展与海量数据承载压力

车队规模扩大后，图像、点云、CAN 信号、语音及标签数据持续并发写入云端。智驾样本的标签维度会随算法策略、车型配置和场景定义持续增加，传统表结构在扩列和高频读写之间难以兼顾，容易影响样本入库、查询和训练节奏。

2. 多源异构数据整合困难

车辆档案、事件元数据、图像点云原始文件及 Embedding 向量，通常分别存放于关系库/文档库、搜索引擎、向量库、缓存、对象存储和数仓等多套系统。跨引擎关联依赖应用层拼接，数据孤岛明显，维护复杂。

3. 复杂场景检索与长尾样本挖掘效率不足

向量数据与业务标签分库存放时，危险工况和 corner case 筛选往往需要"向量召回—关系库过滤—对象存储取文件"的多跳流程。面对不断新增的检索条件，既影响响应速度，也限制高价值长尾场景的发现效率。

三、TDSQL Nexa 解决方案

TDSQL Nexa 通过"核心引擎 + 宽表引擎 + 搜索引擎 + 分析引擎 + AI 计算引擎"的融合架构，结合共享存储，为智能驾驶构建覆盖高性能读写、复杂场景检索、BI 分析和训练数据处理的一体化底座。

1. 宽表引擎与超宽表，支撑动态列和高性能读写

TDSQL Nexa 核心引擎兼容 MySQL 协议，提供统一的事务型数据访问能力；超宽表由宽表引擎承载，面向不断扩展的标签体系支持动态列在线扩展，减少字段变化带来的结构调整负担。宽表引擎以高性能读写能力承接高频样本入库、标签更新和在线查询，保障训练数据持续沉淀与快速调用。

2. 搜索引擎与混合检索，加速 corner case 发现

corner case 的发现主要由 TDSQL Nexa 搜索引擎承担。搜索引擎兼容 OpenSearch 协议，支持向量相似度、关键词、地理位置与结构化标签的 Hybrid Search 混合检索。图像或点云 Embedding 可与天气、道路、车辆状态、接管事件等标签联合过滤，训练团队能够通过一次检索定位符合多重条件的危险工况和高价值长尾样本，避免跨系统多跳拼接。

3. 分析引擎，支撑 BI 报表与标签可视化分析

TDSQL Nexa 分析引擎面向大规模标签数据、样本元数据和业务明细提供 OLAP 分析能力，可直接支撑 BI 报表与可视化看板建设。围绕场景标签、道路类型、天气条件、车辆状态和事件类型等维度，业务团队可进行标签分布、占比、交叉组合、覆盖度及样本规模分析，并支持从聚合结果下钻至样本明细。通过统一的数据入口，训练、产品和运营团队能够围绕同一套标签口径进行可视化分析，减少多系统取数和重复加工。

4. 湖上计算与弹性算力，缩短数据加工路径

AI 计算引擎支持 CPU、GPU 混合弹性算力，可在湖表数据上完成训练前的数据加工、样本宽表构建和特征准备，减少不必要的数据库复制。分析引擎可直接服务仿真结果统计、运营分析及业务报表，高频 SQL 可通过 AI 优化能力提升执行效率。

四、典型应用场景

场景一：智能驾驶训练数据链路

保留现有车端采集、云端接入与流处理体系，将在线存储、检索、数据湖与训练数据处理能力集中至 TDSQL Nexa。宽表引擎以超宽表承接样本元数据与动态标签的高频读写；搜索引擎基于向量与标签混合检索完成 corner case 筛选；AI 计算引擎进一步完成样本加工与特征准备，提升训练数据供给效率。

场景二：车联网数据本体治理

面向车辆档案、事件数据、标签、向量及多媒体文件，构建统一的数据组织与检索视图。宽表引擎提供的超宽表可适配不断演进的业务字段，降低标签新增和数据组织成本；搜索与分析能力协同支撑自驾与座舱等场景的联合查询和数据分析。

场景三：BI 报表与标签可视化分析

基于分析引擎，业务人员可围绕车型、场景、道路、天气、事件和标签等维度搭建自助式 BI 报表及可视化看板，直观洞察样本结构和标签覆盖情况；算法团队可由聚合指标下钻至关联样本明细，快速识别数据薄弱区域，为采集策略调整与训练任务规划提供数据依据。

场景四：仿真回放与故障诊断

分析引擎可直接面向湖表完成大规模统计与明细分析，支撑仿真任务失败案例聚类、故障定位、运营洞察和业务报表生成，缩短从数据发现到分析决策的路径。

五、业务价值总结

TDSQL Nexa 为智能驾驶企业带来以下价值：

- **提升数据读写效率：**通过宽表引擎承载超宽表的动态列扩展和高性能读写能力，适应标签体系持续演进，保障高频样本与标签数据的快速写入和访问；
- **提升样本挖掘效率：**由搜索引擎提供向量与标签混合检索，将 corner case 筛选从多跳拼接简化为一次检索，缩短高价值样本发现路径；
- **降低运维成本：**以一体化多引擎架构收敛关系、宽表、搜索、向量、湖表分析等分散能力，统一管理入口，降低长期维护复杂度；
- **提升业务洞察：**通过 BI 报表和标签可视化分析，帮助业务与算法团队快速掌握样本结构、标签覆盖及数据缺口，为数据运营和训练规划提供依据；
- **降低接入门槛：**兼容 MySQL、OpenSearch 等主流协议，存量应用可渐进迁移，外部数据源可按需接入；
- **支持规模化扩展：**面向高频写入、动态字段、海量标签和弹性算力需求，支撑车队与业务规模持续增长。

TDSQL Nexa 将以统一、高效、可信的多模数据底座，帮助智能驾驶企业加快数据价值释放，推动"采集—检索—训练—仿真—迭代"的业务闭环持续提效。