



StoneAtom

StoneDB

产品白皮书

文档版本：1.1

文档发布日期：2022-10-30

杭州石原子科技有限公司

版权声明

版权所有©2022杭州石原子科技有限公司

本文档的著作权归杭州石原子科技有限公司（以下简称“石原子”）或其许可方所有。未经石原子书面同意，在版权权利限制下，任何单位或个人不得以任何形式全部或部分地复制、摘录、修改、传播本文档内容，或将本文档翻译为其他语言，若发现上述侵权行为，石原子将依法追究行为人的法律责任。

商标声明



StoneAtom
石原子

与其他石原子提供的产品、服务相关的商标均为石原子或其关联方所有。本文档中涉及的第三方的商标，依法由权利人所有。未经石原子（包括其关联方）及有关权利人书面许可，任何单位或个人不得对该等商标进行使用、复制、修改、传播等行为，若发现上述侵权行为，石原子有权依法追究行为人的法律责任。

产品或服务声明

本文档为石原子全部或部分产品或服务的即时情况介绍。由于产品或服务版本升级、调整或其他原因，石原子会不断地更新本文档。您所购买的石原子产品或服务的种类及标准以您与石原子之间的商业合同的约定为准，除非双方另有约定，石原子对本文档内容不做任何明示或默示的承诺或保证。

01/ 概述 05

1.1 市场需求 06

1.2 OLTP 与 OLAP 08

02/ 产品简介 10

03/ 技术架构 11

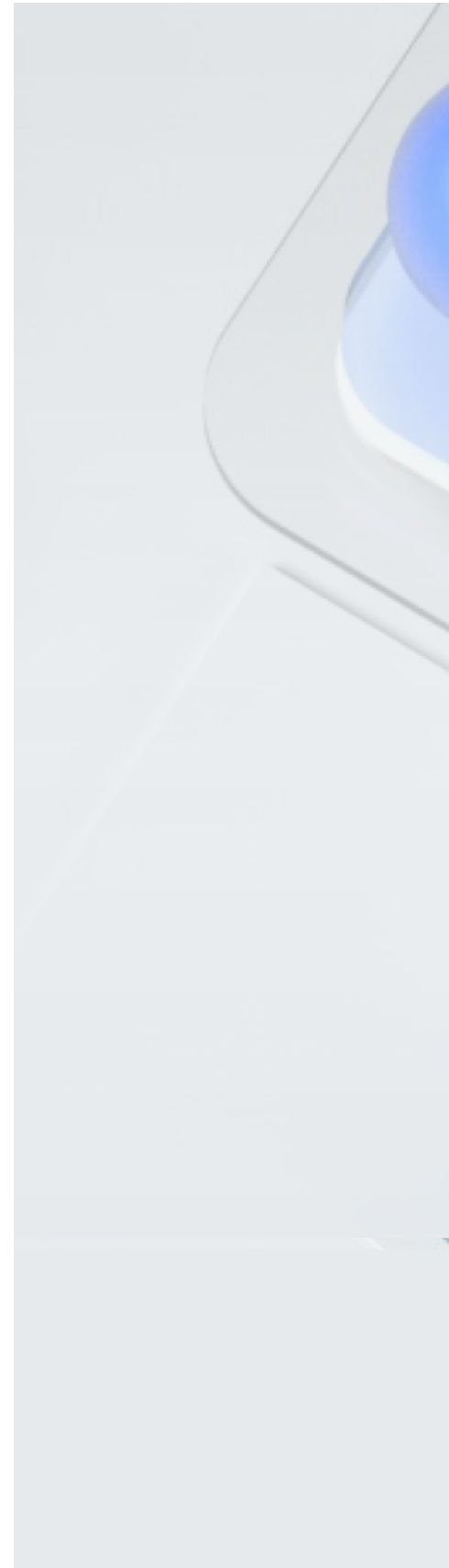
3.1 概述 11

3.2 应用层 12

3.3 服务层 12

3.4 存储引擎层 15

04/ 核心优势 22



05/ 主要功能 28

- 5.1 支持的数据类型 28
- 5.2 DQL 29
- 5.3 DML 30
- 5.4 DDL 30

06/ 提供的服务 31

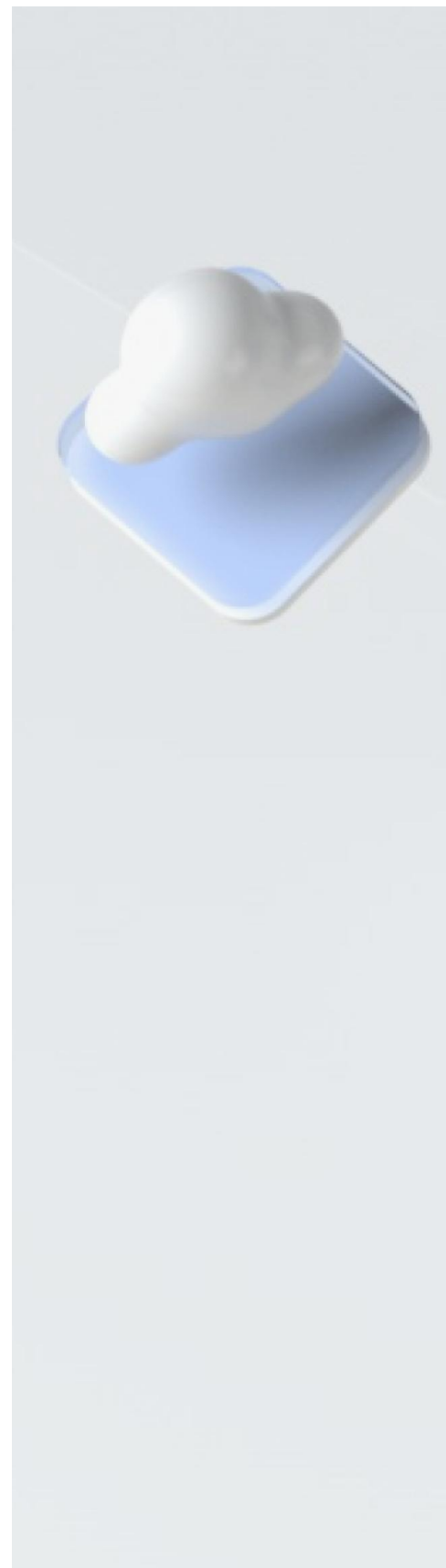
- 6.1 免费服务 31
- 6.2 收费服务 32

07/ 性能对比 33

- 7.1 StoneDB 对比 MySQL 33
 - 7.2 StoneDB 对比 ClickHouse 36
-

08/ 应用场景 37

- 8.1 数字营销场景 37
- 8.2 CRM 系统场景 39
- 8.3 金融风控场景 41
- 8.4 先进制造 IoT 实时场景 43



01 / 概述

数据库是按数据结构来存储和管理数据的计算机软件系统。数据库作为最重要的基础软件，是确保计算机系统稳定运行的基石。自问世之日起，数据库一直是计算的核心，也是企业软件的基础层。企业软件的持久性要求存储的数据在会话结束后立即可用，数据库确保并提供这种持久性。由于更换数据库既困难又昂贵，且往往会对业务运行造成影响，企业管理者在选择数据库时，必须做长久打算，既要考虑总体成本，也要保障企业敏捷性。

1.1 市场需求

HTAP (Hybrid Transaction/Analytical Processing, 混合事务和分析处理) 是指可以同时支持 OLTP (Online Transaction Processing, 在线事务处理) 和 OLAP (Online Analytical Processing, 在线数据分析) 的一种新兴数据库技术。HTAP 概念由权威调研机构 Gartner 在 2014 年的《混合事务/分析处理促进重大商业创新》报告中提出, 并预测 HTAP 数据库将成为数据库领域的重要发展趋势, 一个集成的数据平台将会加速数字化转型。

数据商业价值巨大

许多调研机构都曾发起过数据分析价值的调查。MIT Sloan 报告称受访公司在每 13 美元的投入中就有 1 美元投入到数据分析中; Nucleus Research 的报告显示有 74% 的公司想成为数据驱动型的公司; 福布斯的研究表明, 假设某公司采用更为先进的数据分析手段, 那该公司超越竞争对手的可能性将翻倍。

实时分析的优势

Gartner 的《混合事务/分析处理促进重大商业创新》报告表明, HTAP 数据库比单一的数据库具备更广泛的优势。在传统的应用架构中, 事务系统和分析系统相互独立, 需要通过 ETL (Extract, Transform, Load, 提取、转换和加载) 工具把数据从事务型数据库中导入到分析型数据库再做分析决策。然而, ETL 本身时延较大, 导致数据的商业价值大打折扣。数字业务以及对业务做出实时响应的需求意味着“事后”分析已不再适用。HTAP 数据库支持对实时事务数据进行高级分析, 让企业能够抓住每一次稍纵即逝的商业机会。

开源数据库势头猛进

作为基础软件的数据库在开源趋势下取得了快速发展。DB-Engines 数据显示，2021 年开源数据库的流行指数首次超过商业数据库。截至 2022 年中下旬，DB-Engines 榜单在全球范围内共收录 395 款数据库，其中 206 款为开源数据库，占总数的 52.15%。排名前十的数据库中，开源数据库占据六席。著名调研机构 Ashnik 针对 160 个跨行业和跨地域的企业进行了关于开源趋势和数字化转型的调研。该调研报告指出，在后疫情时代，企业的 IT 领导们在基础设施自动化、云、实时分析以及安全等领域更倾向于使用开源产品，且 74% 的受访企业表示他们在 2022 年会持续并大量增加企业版或社区版开源数据库的使用，59% 的受访企业表示他们将持续增加社区版开源数据库的使用。

MySQL 生态对 OLAP 需求不断增加

MySQL 是世界上最流行的开源数据库之一，根据著名预测性销售及营销情报服务商 Slintel 网站数据统计，截至 2022 年，MySQL 在整个数据库行业的市场占有率达到了 43.04%。随着企业数字化转型的兴起，商业上和技术上都对 MySQL 的在线分析（OLAP）能力提出了更高的要求。然而，MySQL 开源社区的研发方向侧重于其事务处理能力的提升，分析处理能力并非研发关注重点。面对庞大的存量市场和增量需求，为 MySQL 进行 OLAP 赋能存在极大的市场空间。

1.2 OLTP 与 OLAP

自第一个数据库产品面世以来，OLTP 和 OLAP 之间的需求不断分化，这促使了越来越多的 OLTP 和 OLAP 数据库产品问世，分别用于处理事务型需求和分析型需求。OLTP 和 OLAP 数据库之间通过 ETL 工具进行交流，且这种交流主要是将 OLTP 数据库的数据同步到 OLAP 数据库中。

对比项	OLTP	OLAP
用途	实时控制和运行基本业务操作，分批生成报告	计划、解决问题、支持决策、数据价值洞察
查询类型	简单SQL查询	复杂查询
基本操作	INSERT, UPDATE, DELETE	SELECT
响应时间	毫秒级	取决于分析的价值
数据来源	事务	用于分析的事务的汇总数据
数据更新	由用户、应用和接口发起的简短、快速更新	通过 ETL 定期刷新
用户	面向客户的人员、办事员、在线购物者	数据分析师、业务分析师和高级管理人员等知识型员工
数据库设计	规范化数据库，以提高效率	非规范化数据库，以提高分析速度
案例	MySQL、IBM DB2、Oracle Database、SAP、ASE、AWS RDS	IBM Cognos、MicroStrategy、Microsoft Power BI、Oracle 商业智能企业版、Tableau、Google BigQuery、Amazon Redshift

人们普遍认为 OLTP 数据库不擅长 OLAP，OLAP 数据库不擅长 OLTP。为同时满足 OLTP 和 OLAP 需求，企业通常需要部署多个来自于不同供应商的数据库，然后使用另一组供应商提供的 ETL 工具将这些数据库连接起来。绝大多数情况下，这些数据库都使用 SQL 进行访问、管理。因此，这两种数据库的差别是由操作需求造成，而非程序性需求。

对于企业而言，更多的数据库意味着更高的操作、管理以及信息和战略成本。此外，OLTP 数据库与 OLAP 数据库间数据同步的实时性难以保证，导致企业无法基于最新数据进行分析决策。OLTP 与 OLAP 之间相互独立的现状也极大的阻碍了“洞察到行动 (Insight to Action)”的实现。

因此，市场开始探索基于 MySQL 将 OLTP 和 OLAP 进行深度融合的方案。StoneDB 正是这种方案的一次成功尝试。StoneDB 支持本地和云上部署，其提供的 Tianmu 引擎有着极佳的扩展性。StoneDB 采用大规模并行架构，实现了查询性能大幅度提升。针对 400 GB 数据量的分析查询，StoneDB 比 MySQL 快 40 倍，随着数据量的提升，这一性能差异还将继续扩大。

02 / 产品简介

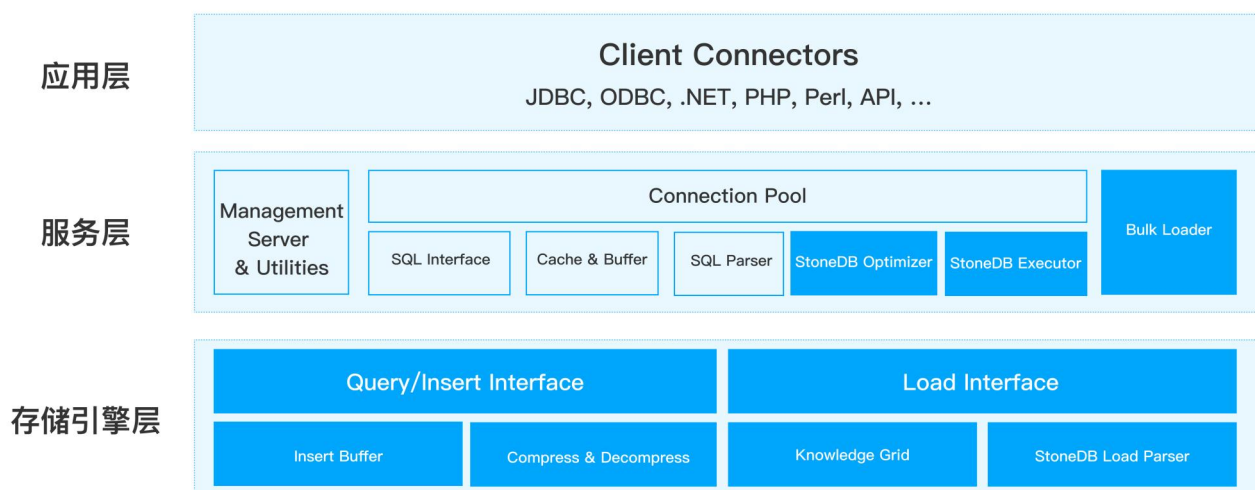
StoneDB 是由石原子公司自主设计、研发的国内首款基于 MySQL 的开源 HTAP 数据库，是一款同时支持在线事务处理与在线分析处理的融合型数据库产品，可实现与 MySQL 数据库的无缝切换。StoneDB 具备超高性能、金融级高可用、实时分析、兼容 MySQL 5.6、MySQL 5.7 协议和 MySQL 生态等重要特性。StoneDB 旨在通过插件化的方式解决 MySQL 数据库本身分析能力欠缺的问题，为用户提供一站式 HTAP 解决方案。

StoneDB 从一开始就设计了向后兼容性，无需任何代码修改，用户即可将业务和数据从 MySQL 迁移至 StoneDB，所有的现有 OLTP 和 OLAP 工作负载都可以继续使用 MySQL 数据库体系运行。

03 / 技术架构

3.1 概述

StoneDB 的系统架构分为应用层、服务层和存储引擎层。StoneDB 沿用了 MySQL 应用层的全部组件以及 MySQL 服务层的部分组件，对 MySQL 服务层进行了改造设计，并设计开发了全新的存储引擎层。



3.2 应用层

应用层是 StoneDB 的系统架构的最上层，沿用了 MySQL 应用层设计，主要提供如下能力：

连接处理

当客户端向服务端发起连接请求后，服务端会从线程池中分配一个线程，这个线程专门负责和这个客户端进行交互。如果客户端断开连接，这个线程也不会被立即销毁，而是被线程池回收，重新分配给新的连接，减少了创建线程和释放线程所花费的时间。

用户鉴权

当客户端连接服务端后，服务端会对发起连接的用户进行鉴权处理，即检查用户名、密码、IP 地址及端口是否正确。如果鉴权失败，则拒绝连接。

安全管理

当客户端成功连接服务端后，服务端会根据权限表里用户的权限来判断用户具体可执行哪些操作。

3.3 服务层

服务层是 StoneDB 的核心层，提供了 StoneDB 数据库系统的所有逻辑功能，包括查询解析、优化、缓存以及所有的内置函数等。该层主要包括如下组件：

Management Server & Utilities（系统管理）

MySQL 原有组件，提供了丰富的数据库管理功能，包括备份与恢复、用户及权限管理和数据库元数据管理。

SQL Interface (SQL 接口)

MySQL 原有组件，用于接收和处理用户的 SQL 语句，得到用户所需要的结果。SQL Interface 支持如下内容：DML (Data Manipulation Language) 语句、DDL (Data Definition Language) 语句、存储过程、视图和触发器。

Caches & Buffers (缓存)

MySQL 原有组件，包括全局和引擎特定的缓存，主要用于提高查询的效率。如果查询缓存中有命中的查询结果，则可以从缓存中直接读取数据，无需再对语句进行解析和执行。该组件由表缓存、记录缓存、key 缓存、权限缓存等一系列缓存组成。

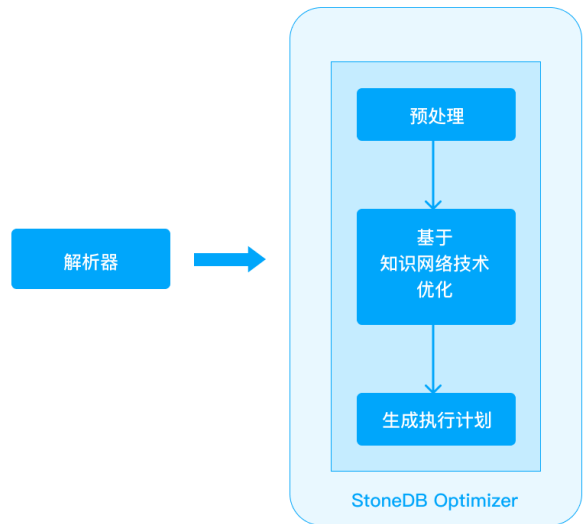
SQL Parser (SQL 解析器)

MySQL 原有组件，主要用于解析查询语句，最终生成语法树。解析步骤如下：

- 1.对查询语句进行语法分析，如果语句语法有错误，则返回相应的错误信息。如果逻辑校验通过，则进入下一步查询缓存环节。
 - 2.查询缓存中是否有对应的语句。如有，则直接读取并返回结果。如无，则进入下一步。
 - 3.对查询语句进行解析，生成语法树。
-

StoneDB Optimizer (查询优化器)

StoneDB 特有组件，用于对传入的 SQL 语句进行优化处理，如表达式转化、子查询转连接等。此外，StoneDB Optimizer 采用了知识网格技术，用网格元数据替代索引，并提供数据过滤、统计信息等，自动生成高级位图 (Character Map)、直方图 (Histogram)，采用高效粗糙集和颗粒计算，形成针对知识网格的查询优化。在选出最优查询路径后，StoneDB Optimizer 调用存储引擎接口，开始执行经过解析和优化之后的 SQL 语句。



StoneDB Executor (StoneDB 执行器)

StoneDB 特有组件，执行器是一个非常重要的组件，SQL 语句经过优化器的优化环节之后，需要通过执行器去调用存储引擎接口，并读取数据返回给客户端。

Bulk Loader (批量加载模块)

StoneDB 特有模块，负责大数据量的批量导入、导出任务，如处理 LOAD DATA INFILE 与 SELECT ... INTO FILE 任务。由于 StoneDB 面向的是海量数据环境，Bulk Loader 作为独立的数据导入模块，并不直接使用 MySQL 的逻辑。此外，该模块支持不同的数据源环境及多语言接入。

3.4 存储引擎层

StoneDB 提供自研的存储引擎，名为 Tianmu。Tianmu 引擎具有高性能、高压缩比等特性，是专门为 AP（Analytical Processing，分析处理）场景设计的列式存储引擎，主要负责数据的存储与提取，向上通过接口与 StoneDB Executor 进行通信，向下与底层文件系统进行交互。Tianmu 引擎的核心组件包括：

Query/Insert Interface（查询/插入接口）

Tianmu 引擎提供了专有的查询接口，用于接收用户的 SQL 语句并进行处理，并使用知识网络格技术优化处理后，转化成执行计划，交由 StoneDB Executor 与存储引擎交互。

Tianmu 引擎也提供了专有的 Insert Interface，负责接收经过 SQL Parser 处理之后的 DML Insert 语句，可通过参数控制其是否经过 Insert Buffer 模块。

Insert Buffer（插入缓冲模块）

StoneDB 的 Insert Buffer 是为整张表的插入做的优化设计。当向表中插入数据时，数据先暂存到 StoneDB 的 Insert Buffer 中，然后再从 Insert Buffer 批量刷新到磁盘中，提高了系统的吞吐量。如果数据不暂存至 Insert Buffer，而是直接写入磁盘，由于 StoneDB 不支持事务，只能往磁盘中逐行写入数据，导致系统吞吐量下降，插入效率低下。StoneDB 的 Insert Buffer 功能默认开启，可通过修改变量 `tianmu_insert_delayed` 的值为 `off` 关闭该功能。

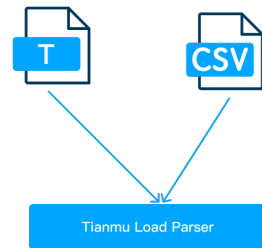
Tianmu 的 Insert Buffer 和 InnoDB 的 Insert Buffer 虽然目的都是提高系统的吞吐量，两者之间还是存在差异。InnoDB 的 Insert Buffer 在接收到数据后会立即刷盘，而 StoneDB 的 Insert Buffer 则不会。因此，在 StoneDB 中，如果在数据插入后立刻针对插入表执行查询，可能查不到完整的数据。

Load Interface（数据导入接口）

Load Interface 主要是接收 `LOAD DATA INFILE` 语句，将请求的文本文件和表信息传入 Load Parser 模块进行解析。

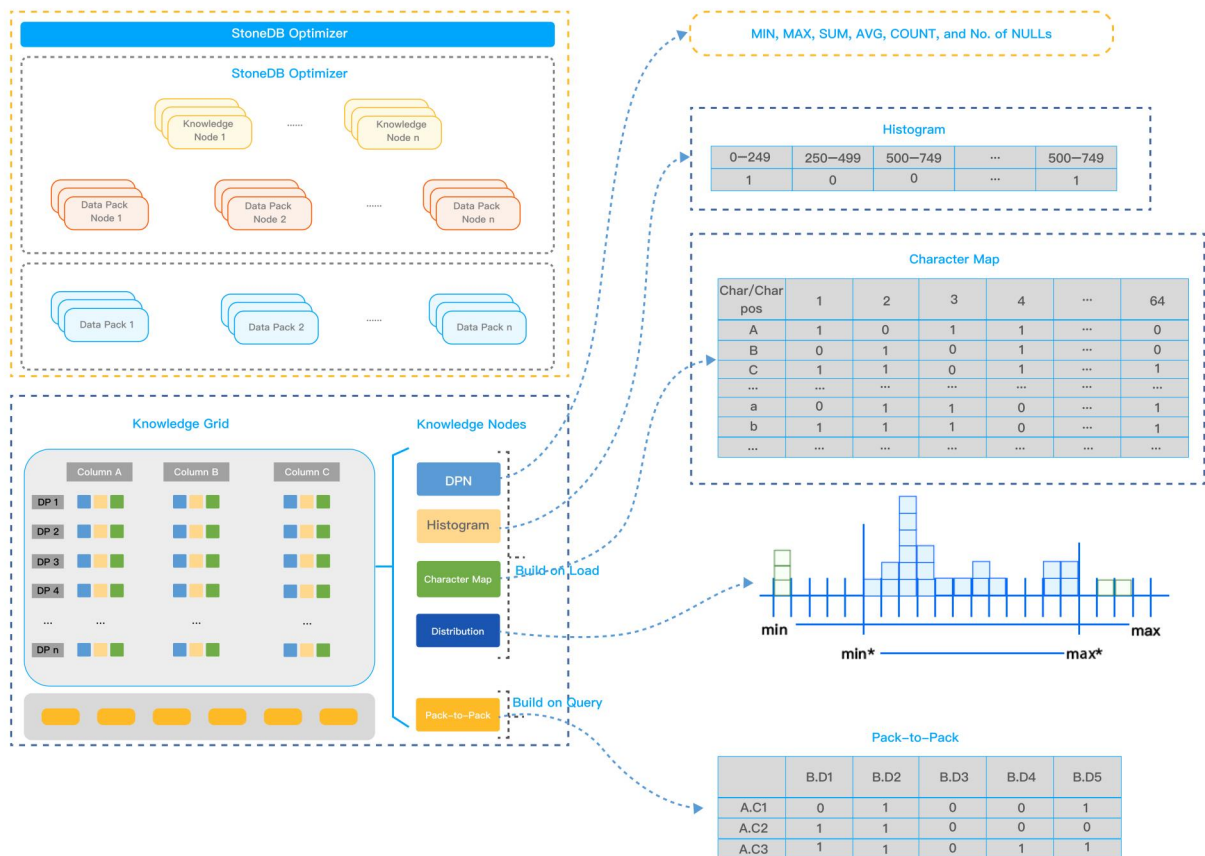
Load Parser（导入解析模块）

Load Parser 模块主要是对数据文本文件或者数据源格式进行处理和解析，并根据固定规则将其转化为可加载入表的数据，Tianmu 引擎的 Load Parser 主要是对 TXT、CSV 等文本文件进行格式解析，并将解析后的数据加载至 Value Cache 中，通过并行插入进程，将数据写入至磁盘。



Knowledge Grid (知识网格模块)

知识网格是由数据包节点 (Data Pack Node) 和知识节点 (Knowledge Node) 组成的。由于数据包 (Data Pack) 是压缩存放的, 读取数据时解压缩成本很高, 因此在查询中如何做到读取尽量少的数据包是提升效率的关键。知识网格正是起到了这样的一个作用, 它能够有效地过滤掉查询中不符合条件的数据, 以最小的代价定位以数据包为最小单元的数据。知识网格的数据大小只占数据总量的 1% 以下, 通常情况下可以加载至内存中, 进一步提升查询效率。对于大部分统计、聚合型查询, StoneDB 往往只需要使用知识网格就能返回查询结果, 这是因为通过知识网格可以消除或大量减少需要解压的数据包, 降低 I/O 消耗, 提高查询响应时间和网络利用率。例如: 数据包节点记录了最大值、最小值、平均值、总和、总记录数、null 值的数量, 如果想对某个列做聚合运算, 那么知识网格就能根据这些元数据很快得到结果, 无需再解压访问底层的数据包。



知识网格涉及如下核心概念：

▶ 知识节点

知识节点除了记录数据包之间或者列之间关系的元数据集合，比如数据包的最小值与最大值范围、列之间的关联关系外，还记录了数据特征以及更深度的统计信息。大部分的知识节点数据是载入数据的时候产生的，另外一部分是查询的时候产生的。

▶ 数据包

数据包用于存放实际数据，是最底层的数据存储单元，每列按照 65536 行切分成一个数据包。每个数据包比列更小，具有更高的压缩比，而每个数据包又比每行更大，具有更好的查询性能。此外，数据包也是知识网格的解压缩单元。

▶ 数据包节点

数据包节点记录了每个数据包中列的最大值、最小值、平均值、总和、总记录数、null 值的数量、压缩方式、占用的字节数。每一个数据包节点对应一个数据包。

► 直方图

数据类型为整数、日期、浮点数的列的统计数据使用直方图（Histogram）表示。在直方图中，一个数据包的最小值到最大值之间被分为 1024 个段，每段占用一个 bit，如果数据包中有实际值处于某一段中，则该段标记为 1，否则标记为 0。直方图在数据加载时自动创建。

下表提供了一个直方图示例。示例说明数据包中有值落在 0~100 和 102301~102400 两个段内。

0-100	101-200	201-300	...	102301-102400
1	0	0	...	1

针对该示例执行如下 SQL 语句：

```
select * from table where id>199 and id<299;
```

根据上述直方图可知，查询没有命中该数据包，因此，当前数据包不满足查询条件，将会直接被丢弃。

▶ 字符映射表

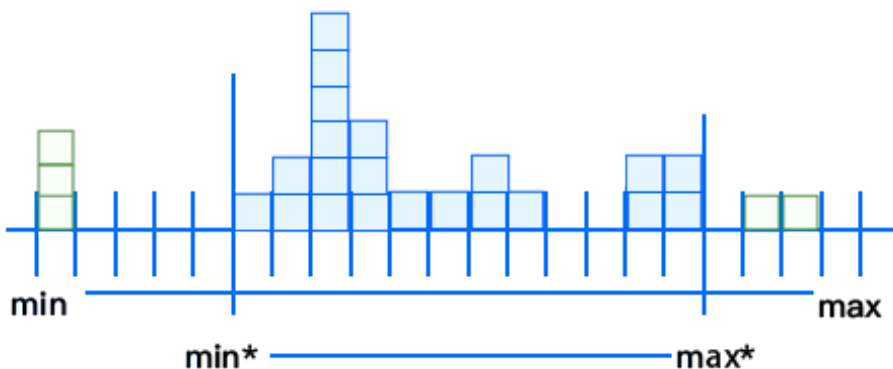
字符映射表 (Character Map) 统计当前数据包内 1~64 长度中 ASCII 字符是否存在。如果存在，对应位置标记为 1，否则标记为 0。字符检索时，按照字符顺序依次对比字符标识值即可知道该数据包是否包含匹配数据。字符映射表在数据被加载时自动创建。

如下的例子中，说明 A 在字符串的第 1 个和第 64 个位置。

字符位置	1	2	...	64
A	1	0	...	1
B	0	1	...	0
C	1	1	...	1
...

▶ 字符映射表

分布统计 (Distributions) 记录当前列中各记录的值分布信息，基于统计信息和粗糙集 (Rough Set) 计算，提供近似查询 (Approximate Query) 支持。分布统计在数据加载时自动创建。



▶ 包对包关系

包对包关系 (Pack-to-Pack) 表示不同表的两个列之间的等值映射表，并以二进制矩阵的形式进行存储，如果符合表关联条件，则标记为1，否则标记为0。包对包关系能帮助在表关联查询的时候快速判断出符合查询条件的数据包，从而提升表关联查询的效率。表关联查询时，包对包关系被自动创建。

如下的例子中，表关联的查询条件是 "A.C=B.D"，在 A.C1 这个数据包中，只有 B.D2 和 B.D5 这两个数据包中有符合表关联条件的值。

	B.D1	B.D2	B.D3	B.D4	B.D5
A.C1	0	1	0	0	1
A.C2	1	1	0	0	0
A.C3	1	1	0	1	1

Compress & Decompress (压缩/解压模块)

• 压缩模块

在 StoneDB 中，数据是按照列模式进行组织的，这种数据组织形式对各类压缩算法十分友好，可依据数据类型选择合适的高效压缩算法。StoneDB 可以支持 20 多种自适应压缩算法，目前主要使用的有 PPM、LZ4、B2、Delta 等。如果列的重复值越高，压缩效果越好。数据压缩后，不仅可以节约存储成本，还可以节约 I/O 和内存。

• 解压缩模块

数据解压缩模块，数据压缩和解压缩的单位是数据包。知识网格技术过滤掉不相关的数据包后，对可疑的数据包需要解压缩才能得到符合查询的结果集。

04 核心优势

完全兼容 MySQL 及 MySQL 生态

StoneDB 完全兼容 MySQL 协议及 MySQL 生态，支持 MySQL 客户端工具和周边生态工具，同时适配各类主流 BI 工具。StoneDB 完整地支持标准 SQL 语法，提供 MySQL 常用功能，包括聚合、关联、排序、分组、子查询、触发器、视图、存储过程等。应用无需进行任何业务代码修改即可从 MySQL 切换至 StoneDB，实现业务和数据的无缝迁移。

最小化硬件开销

StoneDB 对部署的服务器的数量和配置没有要求，可支持单节点最小配置运行，实现了数据库硬件开销最小化，显著降低了 HTAP 系统的部署门槛，也节约了运维成本。而行业内常见的 HTAP 数据库对服务器硬件的数量与配置均有最低要求，相比之下，StoneDB 在使用成本上具有显著优势。

事务 + 分析型混合工作负载

StoneDB 为处理事务型/分析型混合工作负载提供了一种全新的一体化解决方案，实现了事务型负载与分析型负载的有效隔离，解决了事务型工作负载和分析型工作负载资源之间的资源竞争问题。事务数据通过 Binlog 方式实时同步到分析引擎，无需复杂且耗时的 ETL 步骤，不仅满足了事务的 ACID 属性，还解决了业务对数据分析实时性的需求。与传统 TP + AP 架构相比，StoneDB 的系统架构更加简单，提高了数据实时性，并且保证了数据一致性。

实时分析

基于原生 MySQL 复制协议，StoneDB 可将事务数据实时同步至分析引擎，数据更新零延迟，分析引擎可直接分析最新数据，提供最实时的决策支持，实现数据即写即用。StoneDB 大幅简化了实时数据服务的架构复杂度，减少了运维成本，提高了业务敏捷性，让实时分析业务能更快成型落地，加速业务推进速度。

完全开源

StoneDB 的核心代码与相关工具完全开源，严格遵守 GPL 2.0 开源协议。

StoneDB 已建立开源社区，拥有丰富的社区人才，帮助产品快速迭代演进。相关技术文档与课程也已开放，赋能开源社区，使开源社区能对社区用户进行帮助与服务，对用户问题反馈进行及时响应。

活跃的社区生态

StoneDB 构建了一个由用户、开发者、贡献者、StoneDB 爱好者以及众多上下游合作伙伴共同建设的开源社区。社区本着“开源开放、共建共享”的基本原则，围绕“3C（Code、Consumer、Community）”为核心发展战略，为 StoneDB 社区参与者提供丰富的学习内容、专业的技术支持和开发指导、开放的交流和分享平台，与广大数据库相关行业人员共同建设高质量的开源社区。

全面国产化

StoneDB 是一个由国内团队设计开发、自主可控、并拥有完整知识产权的国产开源数据库产品。

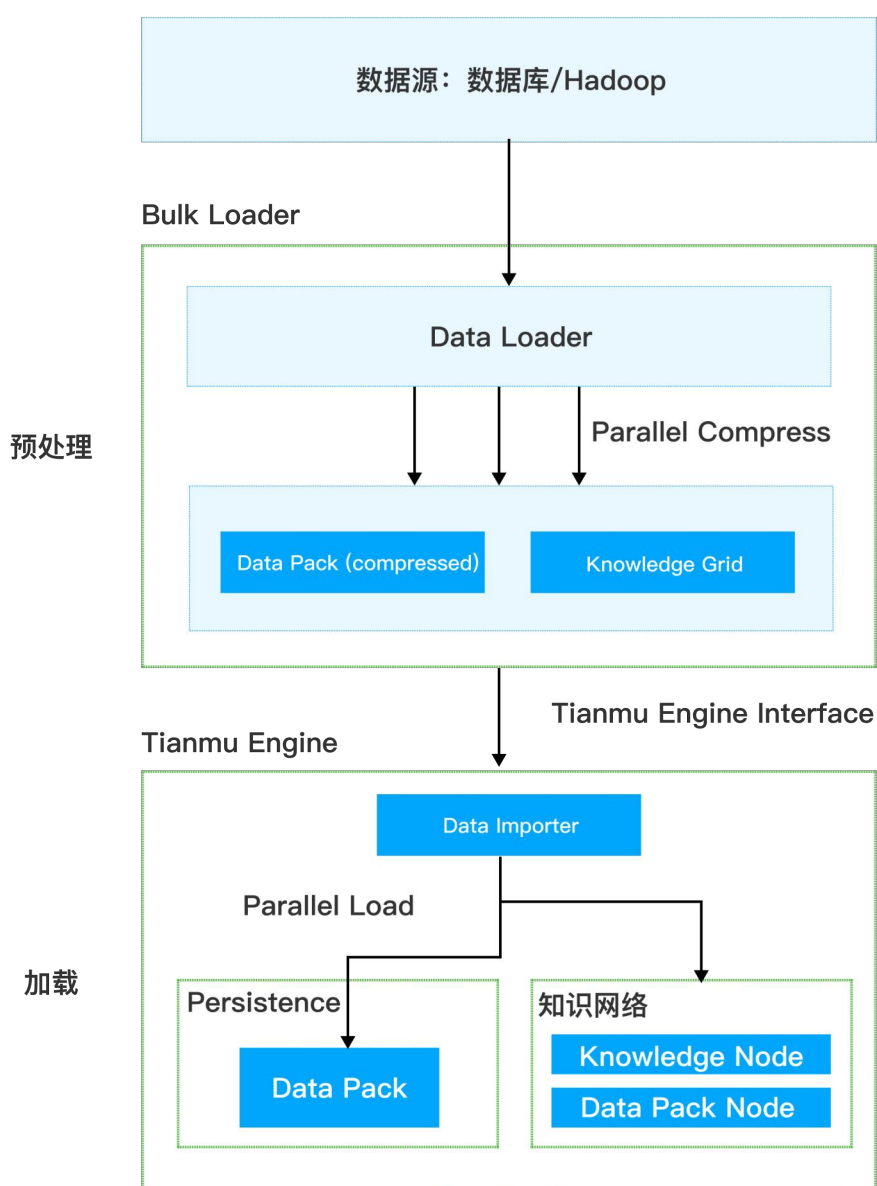
为了助力信创产业，StoneDB 一直坚持进行国产化适配，已经或即将完成鲲鹏、海光、飞腾、龙芯、兆芯等国产处理器适配，以及麒麟、统信等国产化操作系统的适配。未来，StoneDB 还会继续与行业伙伴深度合作，共同发展与完善国产基础软件的产业生态，全面助推国产化建设进程和信创产业应用生态发展，为基础软件国产化的创新发展贡献力量。

列式存储

Tianmu 表在磁盘上按列存储。由于关系型数据库中每一列的数据类型都相同，这种连续的空间存储与行式存储相比，能够实现更高的数据压缩比。在读取数据方面，如果查询只需要读取一个字段，在行式存储中，引擎层向服务层返回的包含该字段的所有行，需要消耗更多的网络带宽和 I/O。而处理同一查询，列式存储只需要返回该字段，极大减少了网络带宽和 I/O 的消耗。此外，列式存储无需再为列创建索引和维护索引。

高性能加载

StoneDB 提供独立的数据导入模块，支持不同的数据源和多语言架构。数据在导入前，StoneDB 会对数据进行预处理，包括数据压缩和知识节点（Knowledge Node）的构建。经过预处理的数据在进入存储引擎后，无需再次执行解析、数据验证以及事务处理等操作。

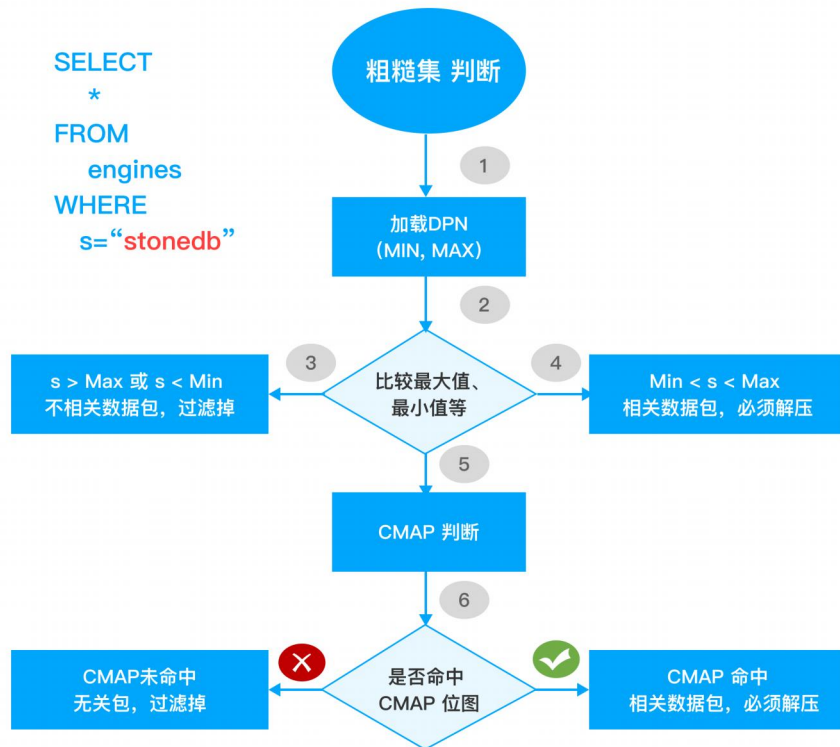


知识网格优化技术

在 StoneDB 中，数据包根据粗糙集概念划分为不相关数据包、可疑数据包、相关数据包：

- 不相关数据包：表示不满足查询条件的数据包，这类数据包直接被忽略。
- 相关数据包：表示满足查询条件的数据包，如果要查询相关的数据包里面的具体数据，需要对数据包进行解压缩，如果根据数据包的元数据节点就能得到数据，那么就不需要解压缩数据包。
- 可疑数据包：表示数据包中的数据部分满足查询条件，需要进一步解压缩数据包才能得到满足条件的数据。





StoneDB 根据知识网格技术过滤掉不相关数据包，对可疑数据包需要进一步解压缩才能得到满足条件的数据。如果能从相关数据包的元数据节点得到结果，则无需对数据包进行解压缩。这样就消除了解压缩数据包的过程和降低 I/O 消耗，提高了查询响应时间和网络利用率。



例如，执行如下 SQL 语句：

```
SELECT count(*) FROM students where score < 550
```

查询结果如下图所示：

	Score	
A	0—200	
B	200—400	
C	400—600	
D	600—800	

通过知识网格可以确定 2 个关联数据包 A 和 B，和 1 个可疑数据包 C。由于此查询包含聚合函数（count），执行逻辑如下：

- 数据包 A 和 B：StoneDB 只需要从其对应的数据包节点中读取 count 值即可，不会消耗 I/O 资源。
 - 数据包 C：StoneDB 读取并解压数据包 C，执行函数计算后，返回结果集。
-

05 / 主要功能

5.1 支持的数据类型

下表列举了部分 StoneDB 支持的数据类型。

分类	数据类型	取值范围
数值	TINYINT	-128~127
	SMALLINT	-32768~32767
	MEDIUMINT	-8388608~8388607
	INT	-2147483647~2147483647
	BIGINT	-9223372036854775808~9223372036854775807
	FLOAT	N/A
	DOUBLE	N/A
	DECIMAL	精度必须小于或等于18
日期	DATE	0001-01-01~9999-12-31
	TIMESTAMP	1970-01-01 08:00:01~2038-01-19 11:14:07
	DATETIME	0001-01-01 00:00:00~9999-12-31 23:59:59
	TIME	-838:59:59~838:59:59
	YEAR	1901~2155

分类	数据类型	取值范围
字符串	CHAR	0~255
	VARCHAR	0~65535
	TINYTEXT	0~255
	TEXT	0~65535
	MEDIUMTEXT	0~16777215
	LONGTEXT	0~4294967295
	TINYBLOB	N/A
	TINYBLOB	N/A
	BLOB	N/A
	MEDIUMBLOB	N/A
	LOB	N/A

5.2 DQL

StoneDB 支持单表或者多表 DQL 以及以下常见查询 SQL：

- UNION/UNION ALL
 - INSTINCT
 - GROUP BY
 - ORDER BY
 - LIMIT (分页查询)
 - INNER JOIN
 - LEFT JOIN
 - RIGHT JOIN
 - LIKE
 - HAVING
 - OR
 - 子查询
 - 重复子查询
 - IN/NOT IN
 - EXISTS/NOT EXISTS
 - CROSS JOIN
-

5.3 DML

StoneDB 支持以下常规 DML 操作：

- INSERT
- UPDATE
- DELETE
- INSERT INTO ... SELECT
- INSERT INTO ... ON DUPLICATE KEY UPDATE
- ...

说明

StoneDB 目前暂不支持 UPDATE 关联子查询、UPDATE 多表关联、REPLACE INTO 等操作。

5.4 DDL

StoneDB 支持以下 DDL 操作：

- CREATE TABLE
 - DROP TABLE
 - TRUNCATE TABLE
 - CREATE TABLE LIKE
 - ALTER TABLE ... ADD COLUMN
 - ALTER TABLE ... DROP COLUMN
 - ALTER TABLE ... ENGINE=innodb
 - ALTER TABLE ... RENAME TO
-

06 提供的服务

6.1 免费服务

StoneDB 的所有功能代码完全开源，可免费下载，用户可自行安装使用。同时石原子官方免费提供各种使用所需的工具及资料，包括：

- 源代码
 - 周边工具
 - 产品白皮书
 - 使用手册
 - 安装部署手册
 - 开源社区
-

6.2 收费服务

为了让用户获得更好的产品使用体验，石原子提供付费服务，让用户能将其精力聚焦于自身业务发展上。服务内容如下表所示：

服务项	服务内容	验收标准
安装部署	StoneDB 及相关周边软件的安装部署服务	软件可正常运行
人员培训	对客户的产品使用人员进行的操作培训	操作人员可以使用产品正常开展工作
专家服务	针对产品使用中的性能提升、疑难杂症等非产品问题处理等提供的现场服务支持	问题解决
运维服务	软件升级、运维、硬件资源扩缩容等	按要求实现软件正常运行
数据迁移	将原有数据迁移至 StoneDB	数据迁移至 StoneDB

服务收费分为专项模式和年包模式两种：

- 专项模式：如果用户可以自行处理绝大部分的任务，仅在部分工作内容上需要石原子的服务支持，推荐选择专项模式。专项模式按照单次服务计费，不同的服务项有不同的计费标准，单次服务的工作量由石原子给与评估，并按照对应服务项的计费标准进行收费。
- 年包模式：如果用户聚焦于自身业务的开发，需要石原子提供整体的打包服务支持，推荐选择年包模式。年包模式中涵盖了所有的收费服务项，按照折扣后的人力计价标准提供打包的整体服务。年包模式额外提供 24 小时在线客服，专人工单追踪服务，旨在为用户提供最佳的服务体验。

用户可针对自己本身的实际情况选择对应的收费模式。

07 性能对比

7.1 StoneDB 对比 MySQL

采用性能测试工具 TPC-H V3.0 版本，测试数据量为 10GB，在相同环境、相同配置、相同数据量、相同测试流程为基础的条件下，按照 TPC-H 的测试要求标准，对 StoneDB 与 MySQL 进行了性能对比测试。

测试环境如下表所示：

产品	StoneDB v5.6.24	MySQL v5.6.24
架构	单节点	单节点
配置	Intel(R) Xeon(R) CPU E5-2683 v4 @ 2.10GHz * 64, 128GB 内存, 4 x 900GB SSD	Intel(R) Xeon(R) CPU E5-2683 v4 @ 2.10GHz * 64, 128GB 内存, 4 x 900GB SSD

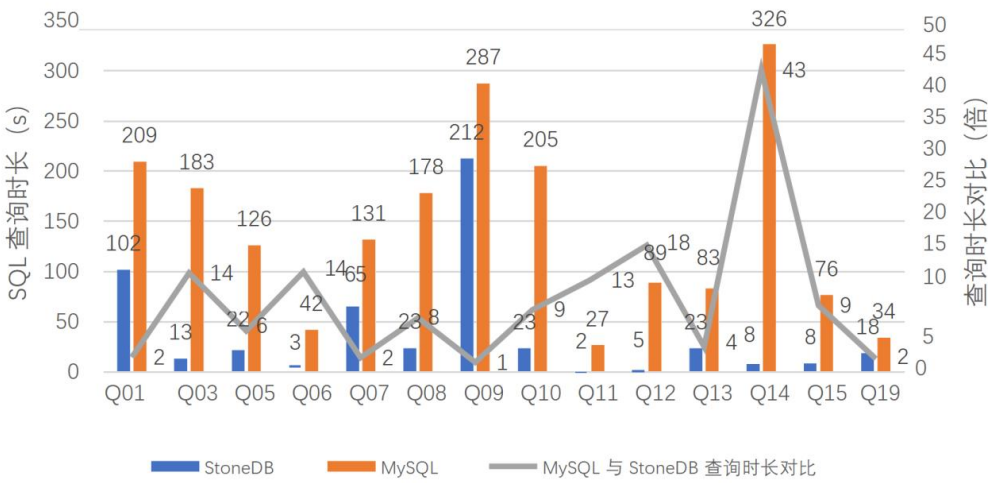
数据规格如下表所示：

表名	数据量
customer	1,500,000
lineitem	59,986,052
nation	25
orders	15,000,000
part	2,000,000
partsupp	8,000,000
region	5
supplier	100,000

根据测试结果，可得出如下结论：

- StoneDB 比 MySQL 查询 SQL 最优条数为 14 条（参见下图），MySQL 相比 StoneDB 的最优 SQL 条数是 5 条。

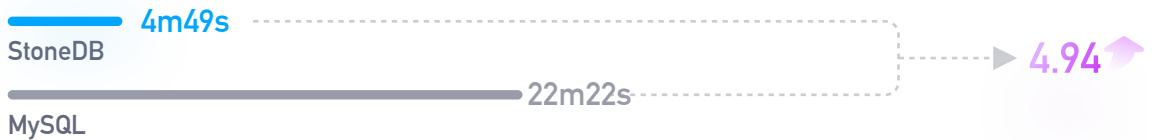
StoneDB 对比 MySQL 查询性能优势



对比维度：占用空间



对比维度：加载时间



7.2 StoneDB 对比 ClickHouse

测试采用性能测试工具 TPC-H V3.0 版本，测试数据量为 100GB，在相同环境、相同配置、相同数据量、相同测试流程为基础的条件下，按照 TPC-H 的测试要求标准，对 StoneDB 与 ClickHouse 进行了性能对比测试。

测试环境如下表所示：

产品	StoneDB v5.7_0.1	ClickHouse v22.5.1.2079
架构	单节点	单节点
配置	Intel(R) Xeon(R) CPU E5-2683 v4 @ 2.10GHz * 64, 256GB 内存	Intel(R) Xeon(R) CPU E5-2683 v4 @ 2.10GHz * 64, 256GB 内存

数据规格如下：

- StoneDB 磁盘占用量：59GB
- ClickHouse 磁盘占用量：42GB

根据测试结果，可得出如下结论：



08 应用场景

8.1 数字营销场景

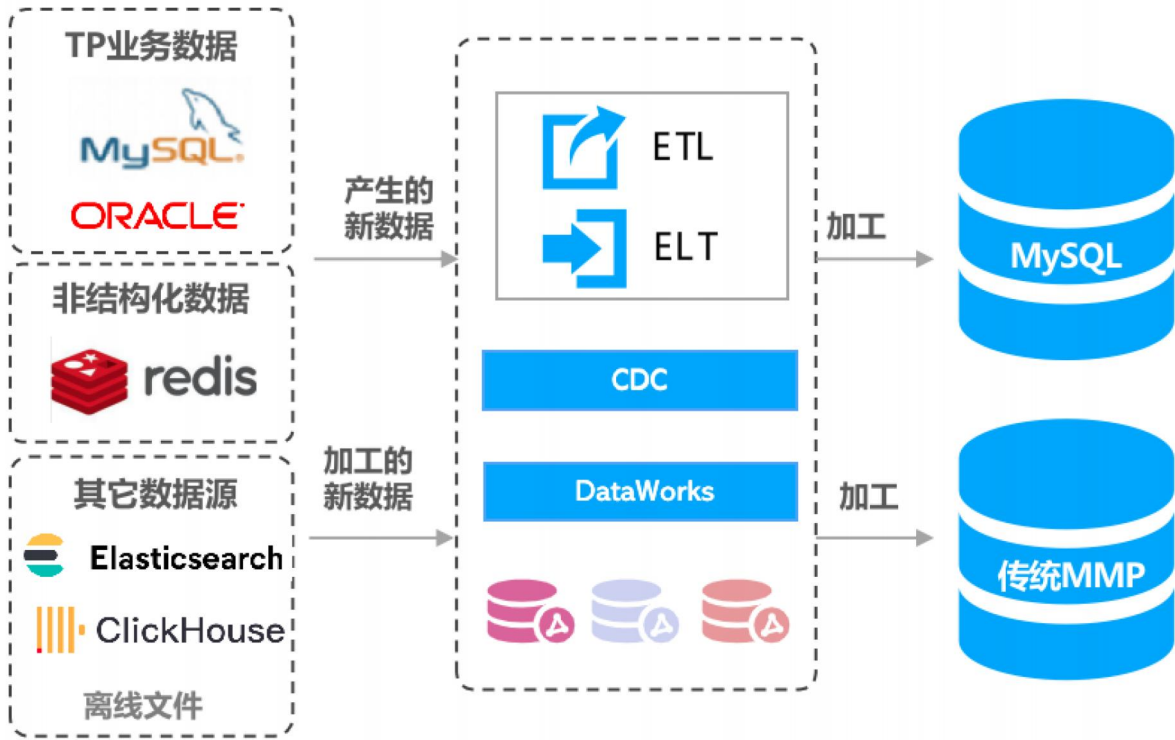
关键字

多源、实时、一站式、低延时

行业痛点

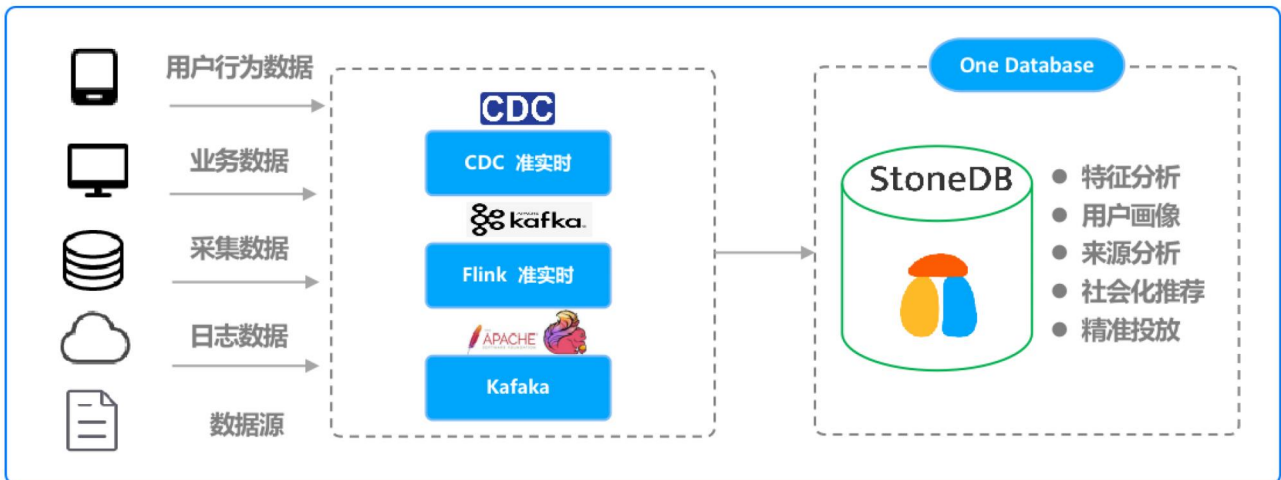
不同于传统“线上+线下”营销模式，数字营销基于大数据驱动，需要实时收集不同客户群体的行为数据，了解客户日常生活消费习惯，动态了解客户当前兴趣倾向、消费偏好和资金往来，同时需要通过对社交裂变的数据追踪，判断社交圈潜在的忠实用户、高价值客户。在具体执行营销业务逻辑前汇总用户行为、用户标签、业务系统交易数据等多源数据。为了追求卓越的用户体验，需要对日志、行为、交易等数据进行流式数据驱动处理，进行去重加工分析，秒级完成整体推荐处理流程。传统架构下的实时数字营销场景在技术上面临如下挑战：

- 架构复杂，数据价值难以快速变现
- 数据分别在不同的数据库中，数据分散形成“数据孤岛”
- 缺乏面向业务场景的高并发查询能力
- 数据需要离线导入进行处理，无法有效支撑在线分析类业务



解决方案

由于数字营销系统数据源的多样性，要求底层数据库可以通过不同的方法进行数据摄入。单一的宽表模型或预聚合模型难以应对复杂的业务逻辑。StoneDB 100% 兼容 MySQL 复制协议，可以原生搭建 MySQL 主从多源复制，也可以通过 Flink CDC 抓取 OLTP 数据库中的数据变更。同时也支持大宽表与多种聚合模型，最大程度满足复杂的业务需求。



方案优势

StoneDB 极致的性能与多样的功能高度适配实时营销业务场景，结合 Flink 构建实时营销系统，用户可以获得以下收益：

- 一栈式解决方案，实现实时业务处理和多维查询的访问隔离，降低技术栈复杂度
- 多源汇聚、实时查询

8.2 CRM 系统场景

关键字

降本、实时、TP + AP、轻量

行业痛点

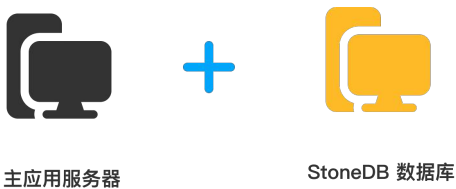
基于“客户全生命周期数字化管理”理念打造的 CRM（Customer Relationship Management，客户关系管理）系统，其主要功能是通过前端营销获客功能与后端的客户管理功能，帮助 B 端客户实现精细化销售管理闭环。主要客户是中小型企业，服务方式分为 SaaS（Software as a Service）服务与私有化部署服务。对于私有化部署的客户，CRM 系统需要为其搭建一套同时满足“TP + AP”使用场景的业务系统，系统要求兼容 MySQL 语法，可以实现 CRM 日常业务的处理，还能进行实时 AP 分析，但传统方案有如下挑战：

- 成本高、培训上手难度大
- 数据库类型多，运维复杂，客户运维投入大
- 查询效率低，无法保证实时性
- 无法兼容 MySQL 语法，业务改动大



解决方案

StoneDB 100% 兼容 MySQL 协议和语法，在同时支持 OLTP 与 OLAP 的混合事务/分析处理（HTAP）能力，可在不牺牲 OLTP 性能和可用性的情况下将分析查询速度提高几个数量级，从而带来业务架构的简化，更少的技术栈，更好的数据实时性。



方案优势

StoneDB 高度适配私有化部署的 CRM 系统场景，用户可以获得以下收益：

- 一体化 HTAP，满足“TP + AP”使用场景
 - 降低了系统实现成本，单节点替代多节点集群，降低了服务器成本
 - 极大降低了运维难度，同时还提升了业务吞吐量，可谓“降本增效”的典型
 - 数据分析实时性极高，数据分析延迟缩短到毫秒级
-

8.3 金融风控场景

关键字

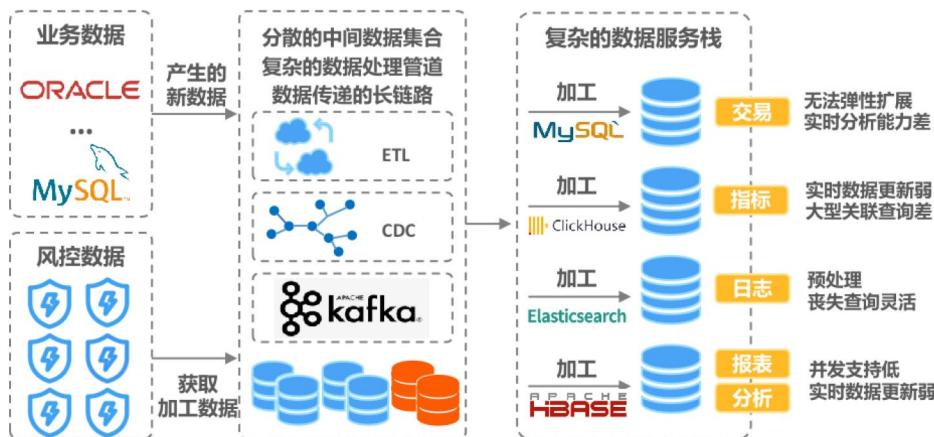
多维度、实时、无 ETL、高效

行业痛点

随着金融需求和生活场景的融合，实时风控的复杂度以及时效性要求随场景服务的发展不断增长，通常需要秒级完成业务流程，并整合决策引擎、征信服务、额度系统、反欺诈系统、公安系统、联网核查、反欺诈等系统能力，整合多方数据源，提供统一的风控能力入口。为了保证数据的实时新鲜，风控业务在数据的导入实时性、查询延迟，模型灵活性上有着极高的要求。

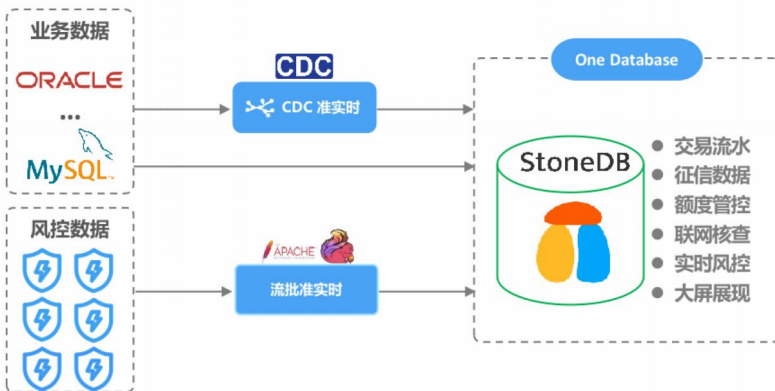
传统方案在实时风控领域面临如下挑战：

- 复杂的数据获取与加工
- 实时价值丧失
- 统一视图构建困难重重
- 复杂数据处理技术栈，数据割裂
- 缺乏灵活性和扩展性



解决方案

聚焦“风险实时预警”小切口，谋划“金融风险处置闭环”的大场景。StoneDB 结合 Kafka 和 Flink 等大数据技术栈，实现将各业务系统最新产生的业务数据、外部监察类数据、本行及三方风控数据实时汇聚到统一的 StoneDB 数据库集群存储管理，提升数据实时获取能力；并进一步对实时汇聚的风控相关数据进行聚合、汇总、分类，变量加工等处理，根据业务的诉求，提供实时风控或实时大屏展现的能力；基于 StoneDB HTAP 架构行存及列存混合处理引擎能力，对业务沉淀和实时汇聚的风控相关数据加工分析处理，形成可用于风控场景消费的价值数据。



方案优势

StoneDB 高效的性能与海量数据存储能力高度适配实时风控业务场景，结合流批准实时技术构建实时风控系统，用户可以获得以下收益：

- 多源结构化数据实时分析
- 简化技术栈
- 统一数据视图
- 无需 ETL
- 提升分析实时性

8.4 先进制造 IoT 实时场景

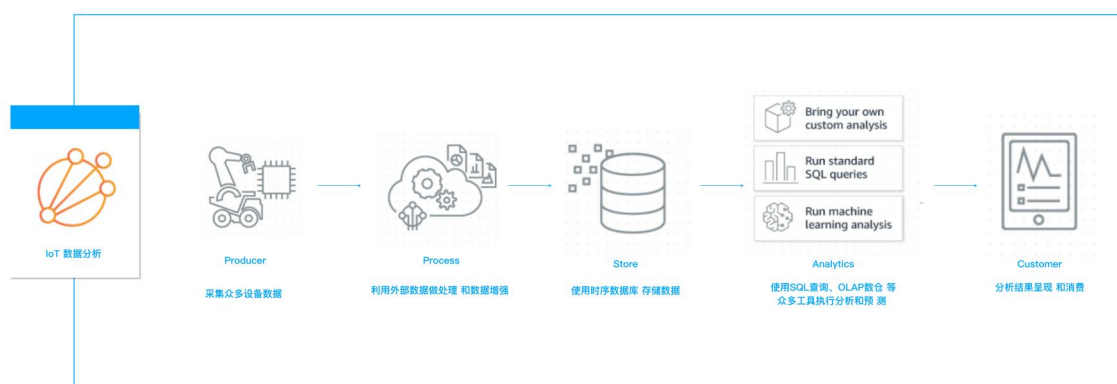
关键字

高效、实时、简洁、缩短

行业痛点

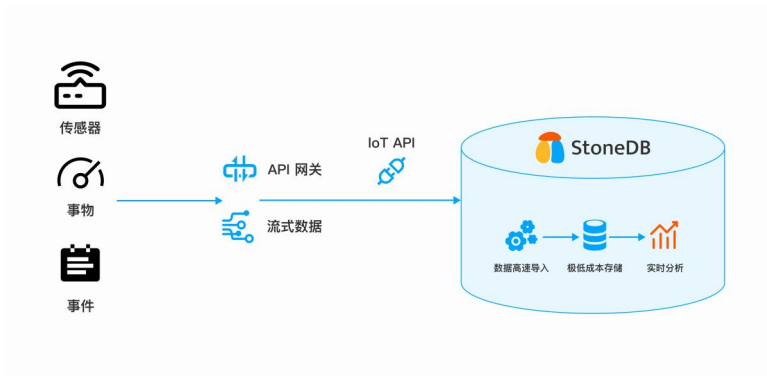
现代制造的背景下，制造工厂更像是一个大型计算系统加上传统的操作工具、大型生产设备的智慧体，每一次生产过程、每一次质量检测、每一个工人劳动量都记录在案。与传统行业不同，物联网行业中的业务没有低峰点，写入量一直都非常大。装配区、高精机加区、结构件区、立库区等功能区域都是智能化、数字化模式的产物。如何积极借助数字化时代的优势，在生产车间导入自动化制造模式，优化运行系统，提升设备生产制造能力，提高其运营的生产力和盈利能力，是企业最为看重的指标。传统方案面临如下挑战：

- 面临无法“高效率、简洁、实时”的数据采集的挑战
- 无法建立预测性维护，无法延长设备寿命和高效运营
- 订单无法进行全程跟踪与分析



解决方案

依据于 StoneDB 可以帮助企业实现：汇聚 ERP（Enterprise Resource Planning，企业资源计划）、MES（Manufacturing Execution System，制造执行系统）、WMS（Warehouse Management System，仓库管理系统）、SRM（Supplier Relationship Management，供应商关系管理）、OMS（Order Management System，订单管理系统）和 PCS（Procedure Control System，过程控制系统）的数据，构成企业的神经系统，可以在 L3~L5 层次之间进行信息的传递，它既能够把业务计划指令传达到生产现场，同时也能将生产现场的信息及时收集、上传和处理。针对每一设备的全链路的路由和时效预测性维护，有效支撑全链路的时效分析与监控，准实时地定位每一设备在每一个环节的状态，延长设备寿命并确保高效运营。



方案优势

StoneDB 海量数据存储能力和高速导入能力高度适配 IoT 实时场景，用户可以获得以下收益：

- 海量流数据实时分析
- 缩短数据传输链路
- 简化技术栈
- 无延迟真实时分析



关于我们

杭州石原子科技有限公司成立于2021年11月，拥有国内顶级的数据库人才，公司专注于企业级云原生实时数仓，并以云中立的数据技术，为全球客户提供低成本、超大规模、高性能的全栈式在线分析服务。通过我们提供世界级的全栈式在线分析服务，以极致弹性、高性价比、极致性能的产品最大化您的数据洞察力，利用数据驱动商业决策，降本增效，推动商业变革。

联系我们

-  官方网站: www.stoneatom.com
-  公司地址: 杭州市余杭区数智引擎创新园B幢2楼
-  业务合作: service@stoneatom.com
-  联系电话: 0571-85220891

