

赋能企业数据 引爆数据价值

石原子科技

数据库AP服务商

致力于给企业提供快速、安全、省钱、省心的数据动力引擎

PART 01

关于我们

About us

- **使命**
Mission 数据价值在线化，省心、省钱、更安全
- **愿景**
Vision 打造云中立的数据世界
- **价值观**
Values 客户第一 结果导向 拥抱变化 团队协作 实事求是

石原子科技是国内领先的**数据库技术**公司。

公司拥有业界开源的MySQL HTAP数据库产品**StoneDB**

企业级云原生实时数仓平台体系**AtomData**。

公司以产品为中心，以客户业务价值为驱动，致力于提供快速、安全、省钱、省心的数据引擎，支撑企业实现数据价值在线化转型突破，以商业洞见、极致性能、成本驱动、搜索驱动数据商业变革。



StoneAtom
石原子

资质与认证： 软件著作权和技术专利

软件著作权
30⁺项

技术专利
10⁺项



资质与认证：行业标准与生态融合



信息技术应用创新产业联盟成员单位
 大数据技术标准推进委员会成员单位
 分布式分析型数据库基础能力专项评测
 中国信通院分布式系统稳定性实验室成员单位
 信息技术应用创新工作委员会技术活动单位
 中国信通院数据库应用创新实验室-汽车行业工作组首批成员



国产化适配

操作系统：欧拉、统信、红旗Linux、中科方德

服务器：华为鲲鹏、曙光

CPU：海光、龙芯、申威、鲲鹏

云厂商：阿里云、华为云、腾讯云、京东云、天翼云

应用层面：四维图新、数澜、半云、华傲等



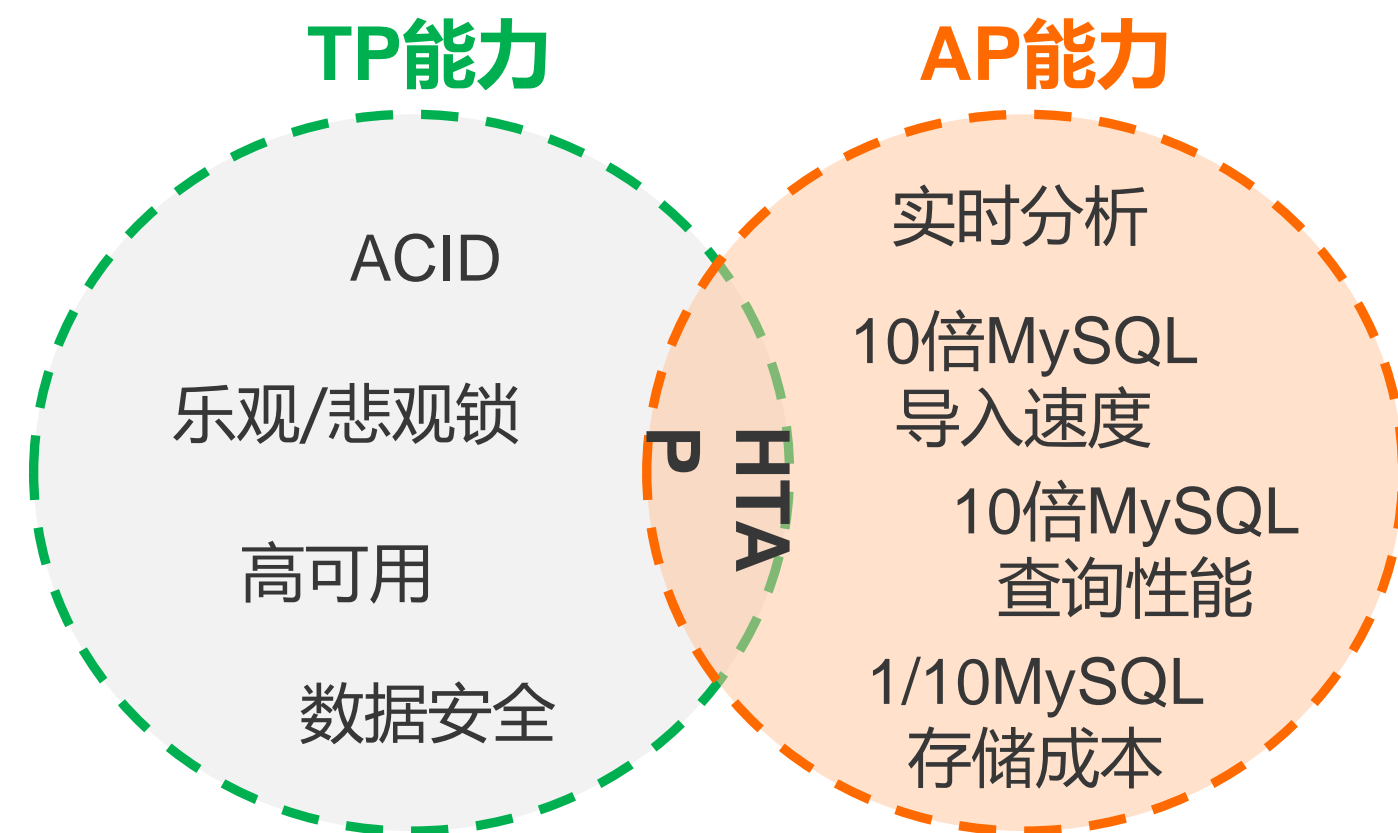


<input type="checkbox"/>	产品名称	首	提交时间	确认通过时间	图谱版本	确认状态	操作
<input type="checkbox"/>	石原子数据库管理系统V1.0	位	2022-08-26 19:04:08	2022-08-29 10:47:54	14	● 已通过	查看 编辑
<input type="checkbox"/>	石原子实时数仓系统 V1.0	位	2022-10-11 09:58:11	2022-10-12 15:26:27		● 已通过	查看 编辑

未来已来---数据分析实时化

小场景：
在业务系统上提供数据分析

大平台：
搭建企业实时数仓平台



业界唯一开源行列混存+内存计算架构的
MySQL**实时HTAP数据库**



新一代**云中立、实时、高性能、低成本**（一站式、全场景、按需计费）的**云原生实时数仓**

PART 02

价值定位- StoneDB

Value Proposition

优势

- 开发部署简单
- 产品成熟稳定
- 开发运维生态丰富

弱点

- 数据容量限制
- 多表连接性能不佳
- 数据分析能力不足

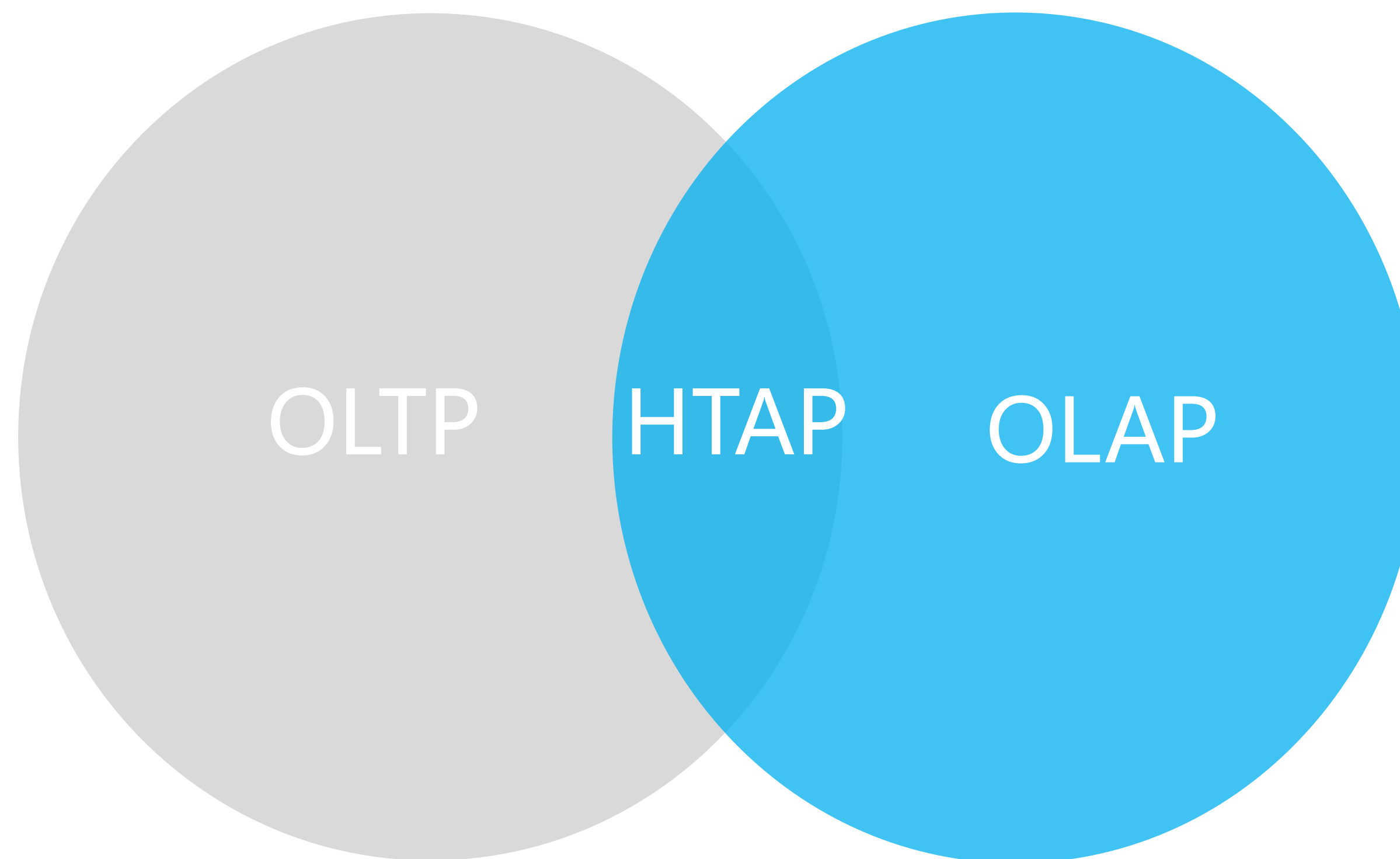
397 systems in ranking, October 2022

Rank			DBMS	Database Model	Score		
Oct 2022	Sep 2022	Oct 2021			Oct 2022	Sep 2022	Oct 2021
1.	1.	1.	Oracle +	Relational, Multi-model ⓘ	1236.37	-1.88	-33.98
2.	2.	2.	MySQL +	Relational, Multi-model ⓘ	1205.38	-7.09	-14.39
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model ⓘ	924.68	-1.62	-45.93
4.	4.	4.	PostgreSQL +	Relational, Multi-model ⓘ	622.72	+2.26	+35.75
5.	5.	5.	MongoDB +	Document, Multi-model ⓘ	486.23	-3.40	-7.32
6.	6.	6.	Redis +	Key-value, Multi-model ⓘ	183.38	+1.91	+12.03
7.	7.	↑8.	Elasticsearch	Search engine, Multi-model ⓘ	151.07	-0.37	-7.19
8.	8.	↓7.	IBM Db2	Relational, Multi-model ⓘ	149.66	-1.73	-16.30
9.	9.	↑11.	Microsoft Access	Relational	138.17	-1.87	+21.79
10.	10.	↓9.	SQLite +	Relational	137.80	-1.02	+8.43

国内首款基于MySQL的开源HTAP数据库

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(Hybrid Transactional and Analytical Processing, 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、MySQL 8.0协议和 MySQL 生态等重要特性。

StoneDB 的目标是通过插件化的方式解决MySQL数据库本身不具备分析能力的问题, 为用户提供一站式 OLTP+OLAP 解决方案。



MySQL生态OLTP业务热数据实时分析场景的**最优解决方案**

HTAP业务场景中MySQL国产化替代的**最佳选择**

StoneDB 用于在不改变原来有架构和应用的前提下支撑 MySQL 无法胜任的分析型业务

MySQL用户面临的困境

- 对数据的分析应用是刚需
- MySQL自身在一定规模数据量下的分析能力有限

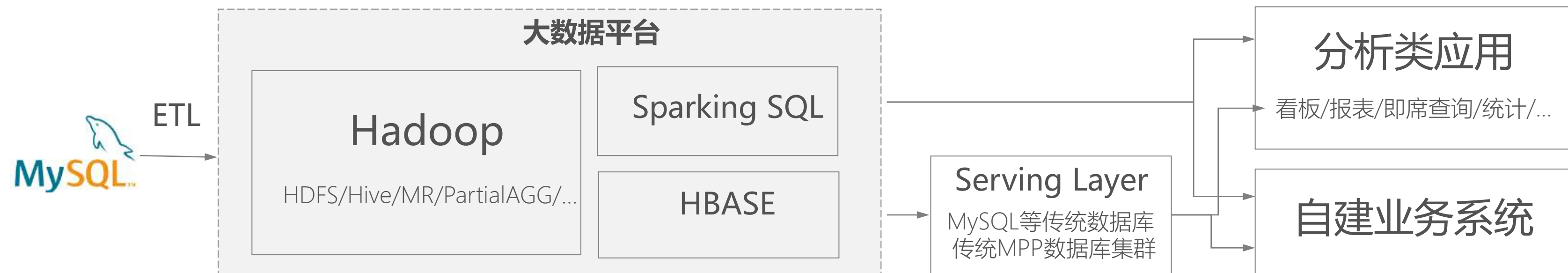
- 单表数据量千万级或亿级
- 多表间关联查询、汇总统计、窗口操作

MySQL响应变慢甚至宕机



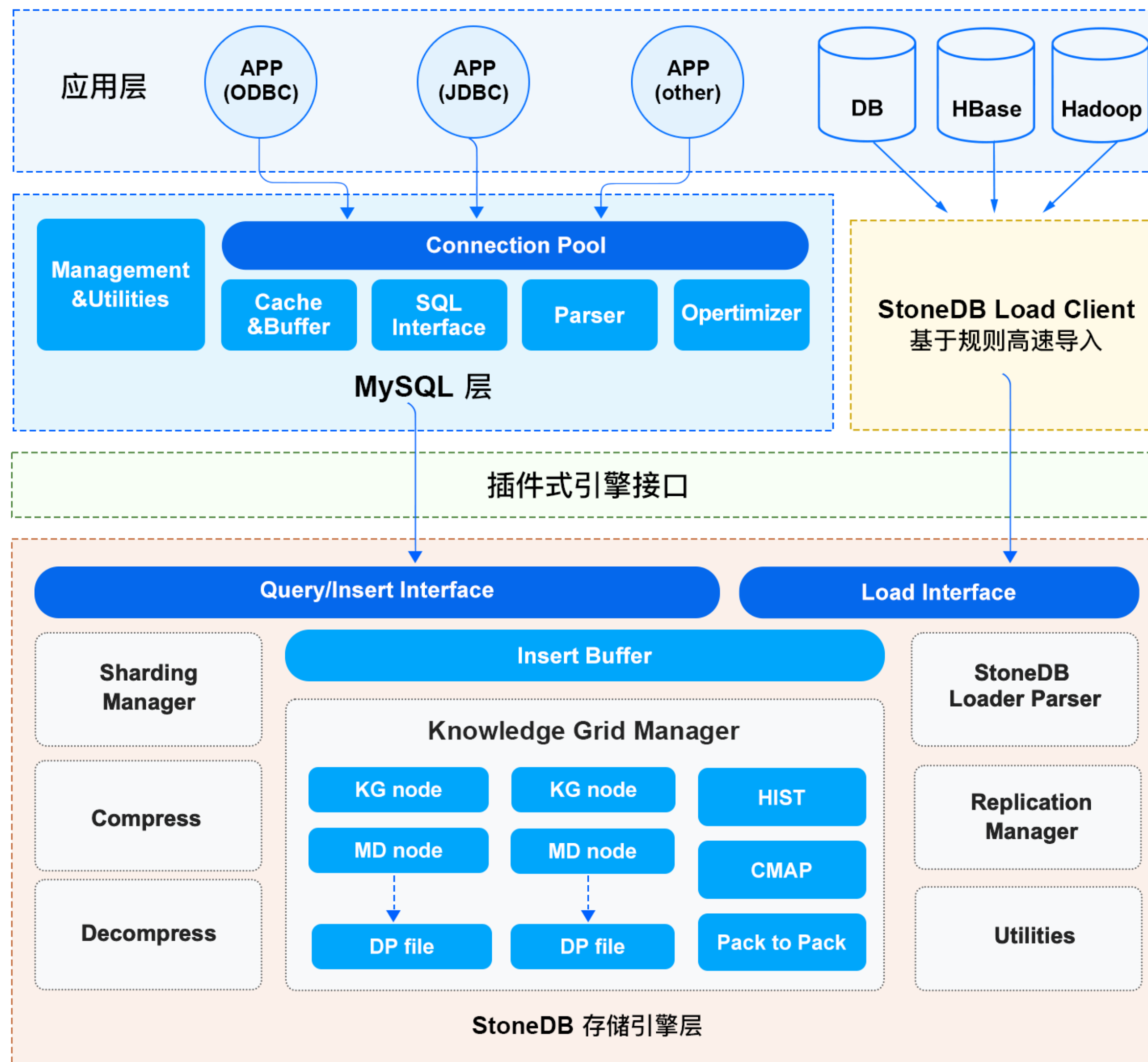
- 不得不借助大数据平台满足分析需要

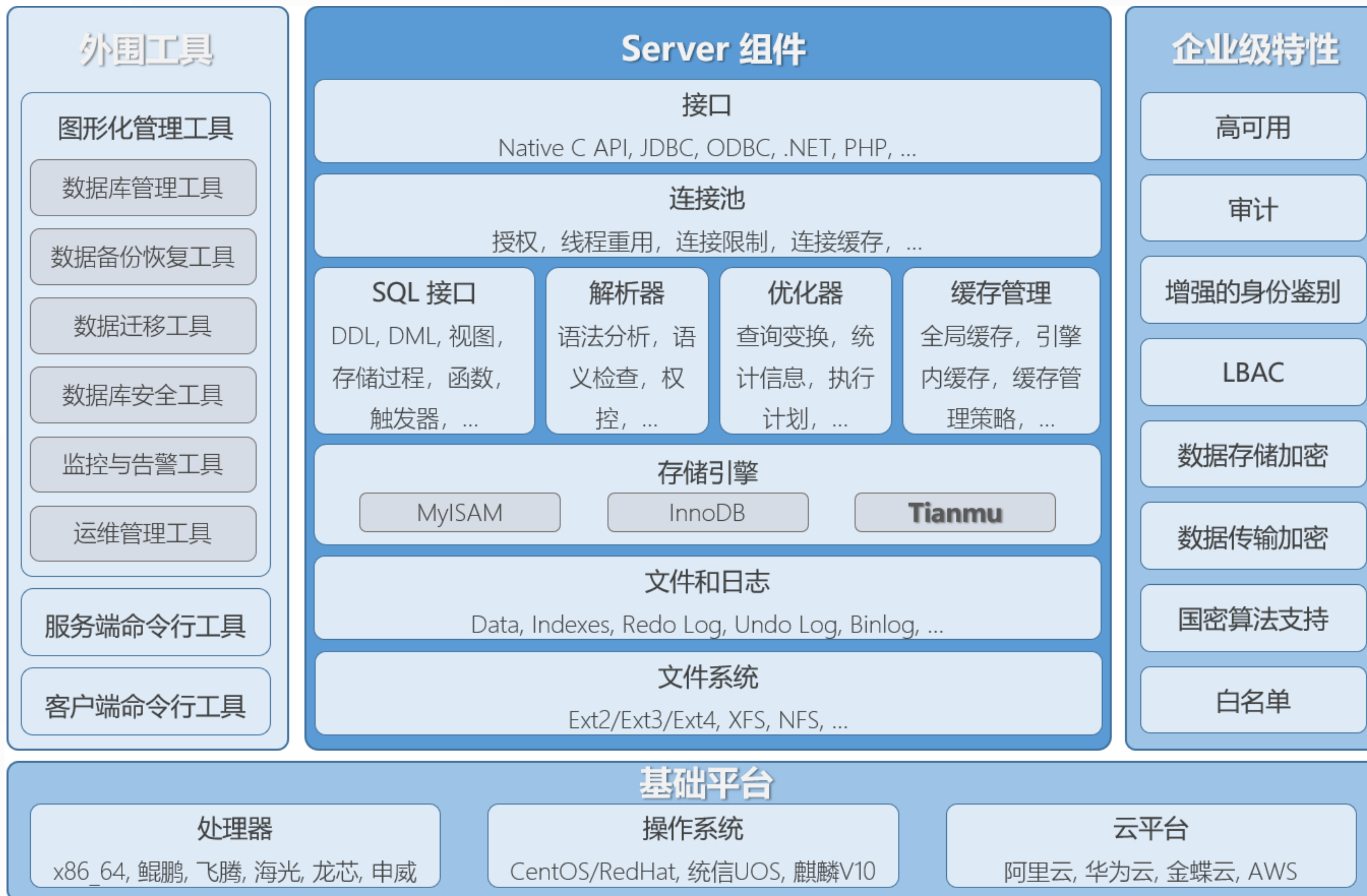
- 结构复杂
- 建设周期长
- 运维成本高
- 实时分析能力不足



StoneDB 可以提供完美解决方案

「行列混存+内存计算」架构, TP + Serving





自研图形化管理界面 StoneDMP



对象区

- TEST1
- audit
- cache
- chenwei
- chenweidb1
- information_schema
- jiangwanqiang
- mysql
- performance_schema
- shiz
- sys
- sys_tianmu
- test
- tianmu_data
- unicom_db
- zz

编辑区

```
1 SELECT * FROM chenwei.user;  
2 SELECT * FROM chenwei.user;  
3 SELECT * FROM chenwei.user;  
4  
5 SELECT *  
6 FROM mysql.db;
```

结果区

✓ 耗时: 5ms SELECT * FROM mysql.db limit 1000

序号	Host	Db	User	Select_priv	Insert_priv
1	localhost	performance_schema	mysql.session	Y	N
2	localhost	sys	mysql.sys	N	N
3	%	audit	audit	Y	Y

审计日志

SQL类型: 全选 数据库名: 全选 用户名: 全选 状态: 全选

操作时间: 开始日期 - 结束日期 近1天 近3天 近1周 近1月

694 条记录

操作时间	操作用户	操作数据库	SQL类型	SQL文本	状态	来源IP
22-09-14 11:14:03					✗失败	
22-09-14 11:14:03	boot		call_procedure	call audit_init_audit;	✗失败	
22-09-14 11:14:18	root		Connect	Connect	✗失败	
22-09-14 11:14:18	root		select	select @@version_comment limit 1	✓成功	
22-09-14 11:14:19	root		Quit	Quit	✗失败	
22-09-14 11:14:32	root		Connect	Connect	✗失败	
22-09-14 11:14:32	root		select	select @@version_comment limit 1	✓成功	

数据库

13 个数据库

名称	字符集	排序规则	创建时间	操作
audit	utf8mb4	utf8mb4_general_ci	-	...
cache	utf8mb4	utf8mb4_general_ci	-	...
chenwei	utf8mb4	utf8mb4_general_ci	-	...
information_schema	utf8	utf8_general_ci	-	...
jiangwanqiang	utf8mb4	utf8mb4_general_ci	-	...
mysql	utf8mb4	utf8mb4_general_ci	-	...
performance_schema	utf8	utf8_general_ci	-	...

information_schema / CHARACTER_SETS

列信息 数据预览 索引 外键 详情 统计信息

前100条数据

序号	CHARACTER_SET_NAME	DEFAULT_COLLATE_NAME	DESCRIPTION	MAXLEN
1	big5	big5_chinese_ci	Big5 Traditional Chinese	2
2	dec8	dec8_swedish_ci	DEC West European	1
3	cp850	cp850_general_ci	DOS West European	1
4	hp8	hp8_english_ci	HP West European	1
5	koi8r	koi8r_general_ci	KOI8-R Relcom Russian	1
6	latin1	latin1_swedish_ci	cp1252 West European	1
7	latin2	latin2_general_ci	ISO 8859-2 Central European	1
8	swe7	swe7_swedish_ci	7bit Swedish	1
9	ascii	ascii_general_ci	US ASCII	1
10	ujis	ujis_japanese_ci	EUC-JP Japanese	3

- 基于MySQL构建, 100%兼容于MySQL生态
- 在语法上和通讯协议上与MySQL保持完全一致

- 自适应压缩技术可以明显节约存储成本
- 无需ETL即可实现MySQL业务热数据的实时分析, 降低系统复杂度
- 不再需要大数据平台

01 完全兼容

02 高性能

03 低成本

04 易用

- 列存和自适应压缩提升了I/O的效能
- 智能索引和知识网格有效提升了数据筛选的效率
- 分析能力相比于MySQL平均提升10~40倍

- MySQL生态中的上下游工具和应用都可以无缝地连接到StoneDB
- 用户的基于MySQL的业务系统无需任何改造就可以迁移到StoneDB

列式存储和数据压缩技术

- 物理列存储架构对数据进行持久化存储
- 内置了多种智能压缩算法

内存计算技术

- 基于业界领先的内存计算原理与架构
- 内存数据按列和按需组织，更加适合承载分析型运算任务

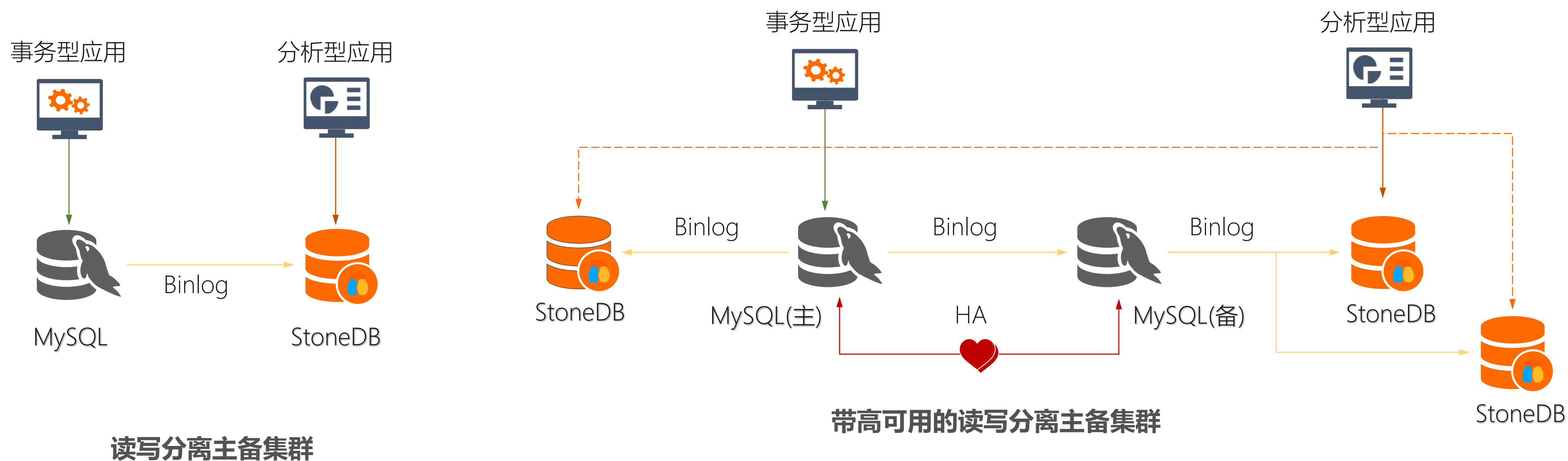
关键技术

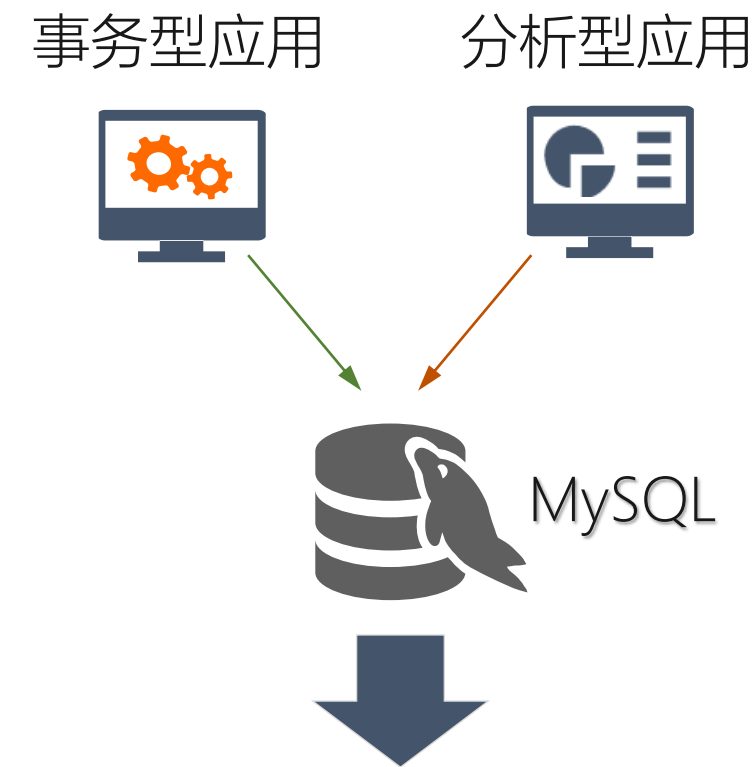
智能知识网格

- 对于经常使用的表间关联查询，自动地基于表间关联关系为关联查询创建智能索引
- 所有智能索引构成知识网格，提供海量数据场景下高效查询分析的技术保障

HTAP数据库StoneDB: 典型部署

- 可以完全替代MySQL (当作MySQL来使用)
- 支持线下部署、云环境私有化部署、阿里云计算巢部署
- 最佳实践**: 在生产环境中, 部署为MySQL主从架构的备机, 响应分析型请求



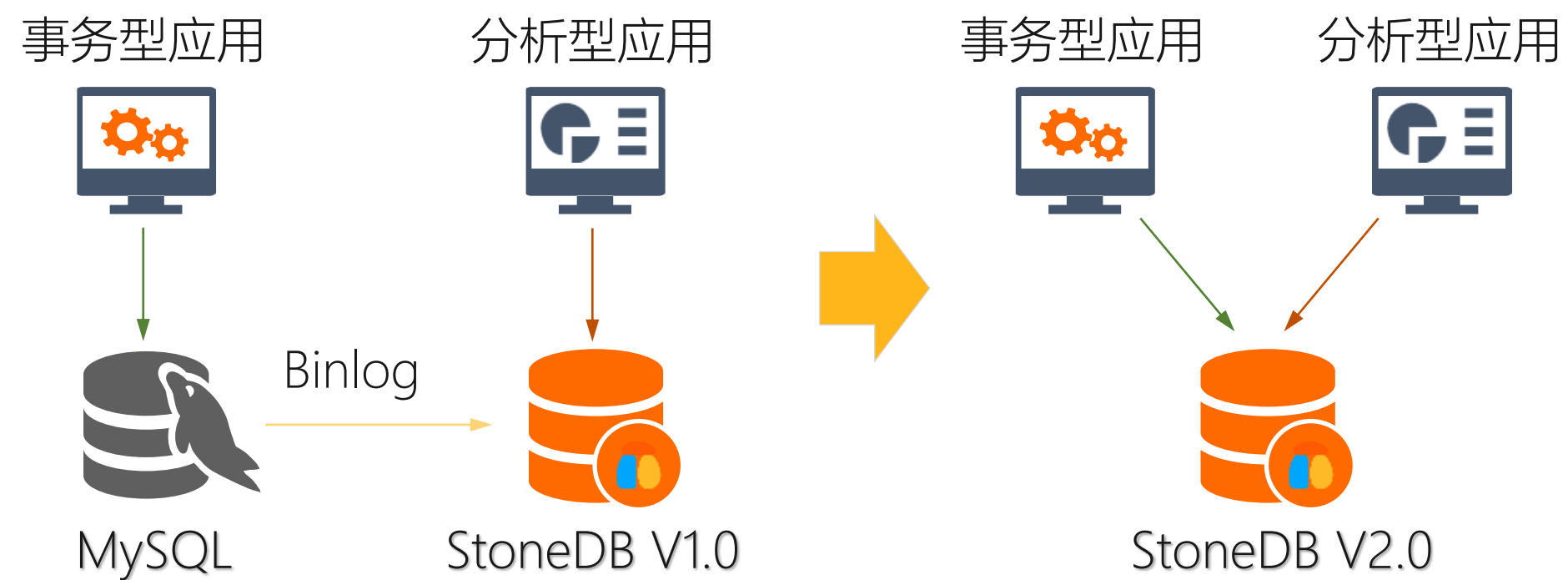


MySQL 原架构

- 事务型应用和分析型应用都由MySQL完成
- 数据量足够大时，分析能力明显下降
- 无法应对复杂分析
- 分析运算会严重影响事务处理的性能，甚至会造成MySQL假死或宕机

StoneDB V1.0 解决方案

- 以备机角色支撑原来架构中的分析型应用
- 完全兼容MySQL语言和生态
- 通过Binlog实现MySQL与StoneDB的数据同步
- 原有分析型业务连接到StoneDB
- 支持通过读写分离组件把分析型请求路由到StoneDB的方式



StoneDB V2.0 解决方案

- 云原生分布式架构，支持弹性扩缩容
- StoneDB 完全替换MySQL
- 一体化行列混存+内存计算架构
- 存算分离，计算能力MPP扩展
- 智能的基于代价的HATP查询引擎
- 原有系统架构和部署完全保持不变
- Native 自动驾驶

行业案例：赋能某CRM厂商降本提质增效



石原子StoneDB协助该企业实现核心业务系统平滑迁移
打造了一体化、智能化的产销体系，满足不断增长的客户定制化需求



▶ 业务快速响应

StoneDB基于HTAP引擎架构，拥有算子下推和并行查询能力，可实现百万级QPS吞吐，流量洪峰场景下性能依旧稳定，轻松应对洪峰期计算压力；同时支持故障秒级恢复，满足核心业务系统对数据可靠性和可用性的需求。

▶ 应用 0 改造

原核心业务系统采用的MySQL+ES，灵活性较差，StoneDB完全兼容MySQL，无需分库分表和应用改造，极大节省了改造成本，同时便于用户拥抱开放生态，增加业务的灵活性；而且性能较开源数据库提升了10倍，数据库基础运维效率和性能大幅提升。



▶ 平滑在线迁移

石原子采用数据在线迁移服务，解决了新老业务之间、各个微服务之间的数据孤岛问题；同时提供数据对比功能，直观展示迁移过程中源库和目标库的数据一致性情况。

数据库访问压力大，作为生产业务系统的核心，可靠性和兼容性要求高、挑战大



数据库访问压力大

核心业务模块灵活定制化后，从一个流程一个业务点，到一个组织一个业务体系，数据量暴增，数据库压力也随之变大。



极致可靠性要求

流程及数据系统属于CRM最核心的生产业务系统，绝对不能接受数据丢失，故障恢复需要达到秒级，可靠性和可用性是核心诉求。



兼容性要求

核心系统的数据库采用MySQL+ES的方案，需要“HTAP架构”作为亮点，因此需要一款完全兼容MySQL，同时提供强大AP能力的数据库。



性能提升：基于StoneDB的极致性能和极速弹性扩容能力，在用户高并发场景下，核心业务及数据库性能依旧稳定，吞吐量**提升68%**，超客户**4倍**预期。



降低改造成本：StoneDB完全兼容开源MySQL，同时提供了强大的列式存储引擎，无需分库分表和应用改造，**节省成本52%**。



打破数据孤岛：StoneDB解决了该企业新老CRM业务模块之间、各个微服务之间的数据孤岛问题，充分调动了各服务之间的联动性，让原先孤立的数据活了起来。

行业案例：家电快消营销分析

客户场景

- **多渠道多系统**：来自不同渠道的营销数据分散在多个系统，数据碎片化，需要整合后进行数据分析
- **多源数据汇聚**：各种渠道的均采用Mysql 数据库，mysql数据库存在短板，数据分析能力不足

客户痛点

- **数据汇聚**：大量不同渠道的营销数据，现有系统无法支持统计分析，大的查询会导致系统崩溃；多个不同营销渠道，跨地域，跨事业部营销数据无法统一分析
- **IT投入成本**：现有系统改造需要额外成本，引入某分布式数据库X，但是X需要多台高端机器搭建集群，投资较大

价值提升

- **时效性提升**：分析性能从T+1→分钟级
- **成本节降**：极高性价比，由10台→1台单机即可完成分析
- **时效提升**：事后复盘→事中实时跟进，即实现跨渠道，跨地区的营销方案实时分析，及时调整市场策略，提升产品销售
- **数据无需迁移**：实施成本降低90%

性能提升
10-100

成本降低
80%

应用代码
0 修改

家电销售场景：销售决策中感受实时数据流动之美

业务背景

- 业务高速发展，数据处理能力面临挑战，中国区迫切需要实时高效获取“新鲜销售数据”；
- 市场与营销多场景的灵活数据计算能力，如经销商返利等；
- 动态的销售数据加工与更新
- 上游系统十几个MySQL实例，数据规模大小约5TB

方案优势

- **数据实时汇聚**：将业务数据和查询流量（接口、在线报表、数据导出）迁移到 StoneDB，通过列存引擎支撑业务即席分析；
- **提升系统性能和连续性**：基于业务库&后备库的架构，提升整个业务系统的吞吐能力、扩展能力以及可用能力
- **需求得以充分满足**：数据多实例汇聚到 StoneDB，为业务提供统一视图，降低跨库、跨业务查询、聚合、消费的实现难度



PART 03

价值定位- AtomData

Value Proposition

数据分析面临新挑战

数据散乱不一致

5+

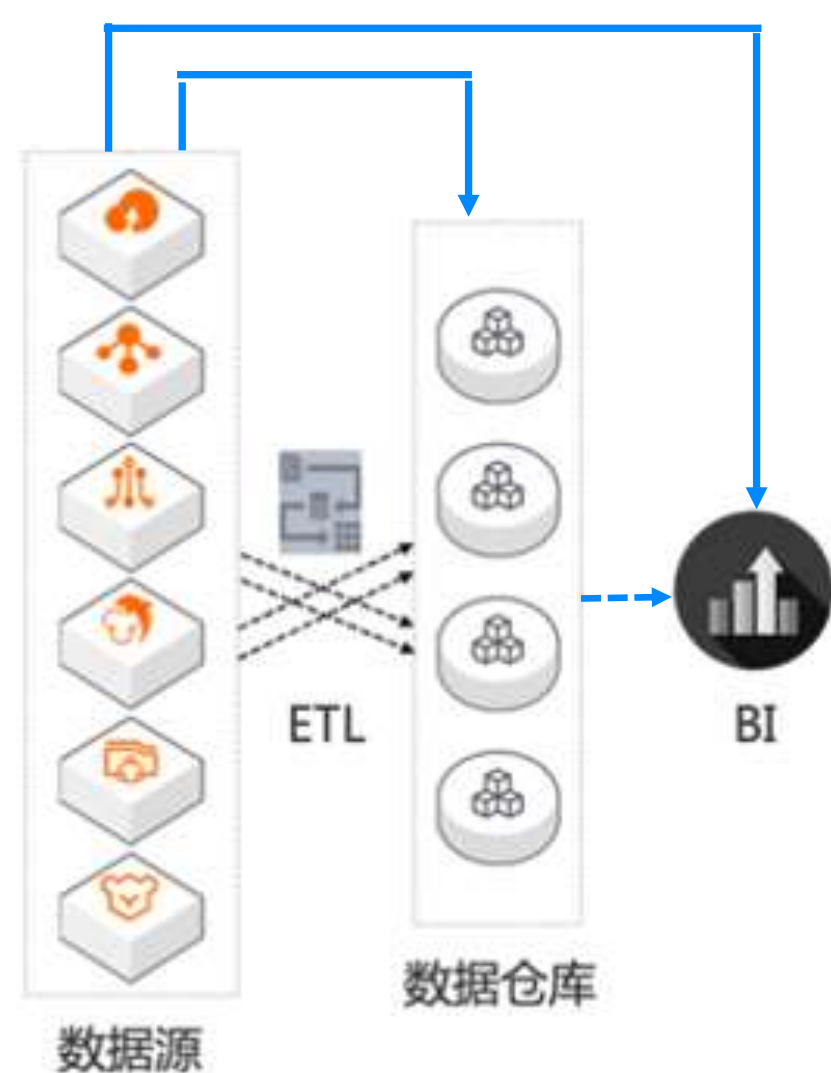
平均数据源



分析不实时

T+1

传统大数据架构



系统极复杂

40+

系统或组件



学习使用成本高

15K+

Hadoop/Spark 月薪

41,951
Hadoop Questions

59,445
Spark Questions

Spark top tags
Scala/PySpark/..

Hadoop top tags
Mapreduce/Hive/..

强一致

实时

一体化

自运维

云原生数仓「一站式」

云数仓走向实时化



扩展性差，扩容成本高
GB级
结构化数据

软硬件一体化



开源框架，时延高
TB级、PB级
结构化、半结构化数据

分布式系统



未完全解耦
TB级、PB级
结构化数据

云化数仓



不支持细颗粒度权限管控
TB级、PB级
结构化、半结构化数据

近实时数仓



离在线一体化、融合分析
PB级
结构化、半结构化、非结构化数据；实时数据

一站式实时数仓

- 100%兼容SQL标准
- 存算分离，按需付费
- 毫秒级读写延迟
- 自动弹性扩缩容
- 高可靠、高可用、低成本
- 上下游生态对接
- 隐私和数据安全

本地化部署 (On Premise)

云化部署 (On Cloud)

AtomData产品定位

离线数仓



星环科技TDH

- 数据管控能力强
- 不支持非结构化数据
- 架构复杂，灵活度低，实施不敏捷
- 组件复杂，成本高
- 计算为T+1模式

OLAP



- 弹性伸缩能力不强
- 抽象模型复杂度高
- 业务分析响应速度慢
- 交互式分析能力数据分析比较弱



AtomData



大数据与数据库一体化

- 轻内核，近乎无状态
- 存算分离
- 高弹性集群，完全负载隔离
- 毫秒级别的读写延迟
- 利用Mysql存储检索元数据
- TB到PB级数据的复杂分析

AtomData 技术架构

• RC节点

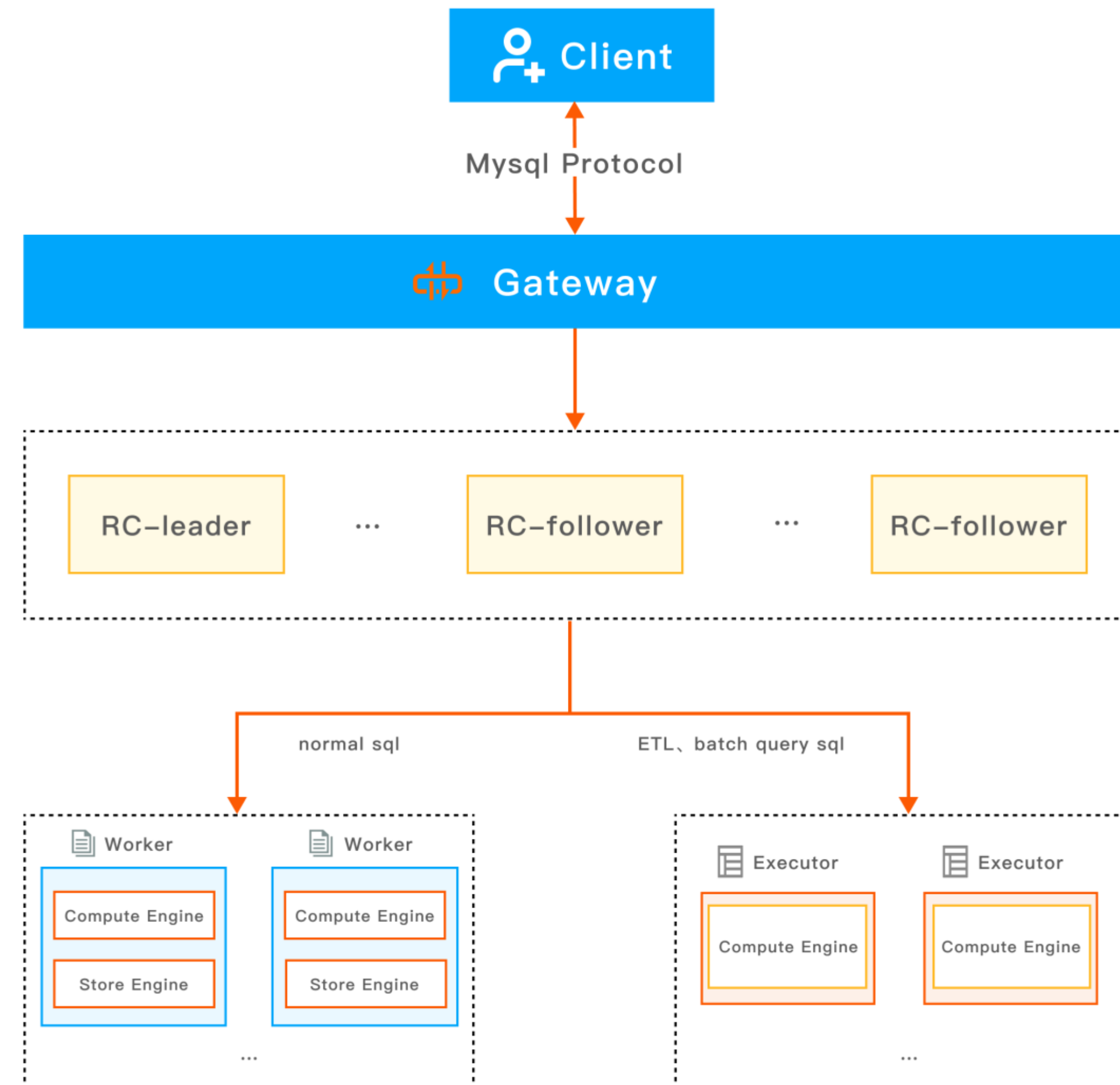
- Resource Controller,控制节点，负责管理客户端连接、SQL路由、SQL解析、生成执行计划、查询优化、资源管理、任务调度等工作。
- RC节点支持容灾部署，同时支持多主架构

• Worker:存算节点

- 内置计算引擎和存储引擎，既负责存储数据也可以参与查询计算。
- Worker可以部署多副本模式保证数据高可用性，副本间通过Raft协议保证数据强一致性。

• Executor:计算节点，

- 内置计算引擎，只负责查询计算。Executor属于无状态节点，支持动态启停。
- 对于计算密集型业务，可以分配额外的计算节点，缓解资源争用



AtomData 的三大优势

明显的性能优势

- 多表复杂分析处理能力强
- 性能表现优异
 - ClickHouse: 7.14倍
 - ADB for MySQL: 1.54倍
 - Doris: 1.52倍
 - GreenPlum: 2.86倍

性价比高

- 整体结构紧凑、无需过多的外部组件即可完成大数据分析
- 具备流批一体的处理能力
- 数据处理链路短
- 数仓建设周期短, 初始拥有成本低
- 产品易于维护, 综合运维成本低

易用

- 100%兼容SQL标准, 所有数据库工具可以直接访问
- 对接多种数据源
- 支持多模数据
- 上下游生态无缝对接
 - 常见的ETL、BI产品
 - 常见的展示分析工具

新一代云中立、实时、高性能、低成本云原生实时数仓

30%~80%

数据分析成本减少



20+BI工具



OLTP+NoSQL 无缝实时同步

**MPP+DAG融合
在离线一体化**

CBO+CodeGen+混合负载

AI扩展+向量检索

结构化+非结构化数据联合分析

GB~100PB+

支持存储扩展



AtomData



安全可靠
数据传输企业级加密

99.99%

高可用性

3~2000节点

支持计算节点扩展

一套存储支持 多场景

写入、点查、多维分析；行列混存+RAFT复制；智能索引

完备的企业级特性

备份/FlashBack/回收站/审计/自建账号/VPC

新一代列式存储

- 物理列存储，内存中组织为行存形式
- 列存-高效的数据查询和复杂分析处理能力
- 行存-快速的数据写出能力

多部署架构

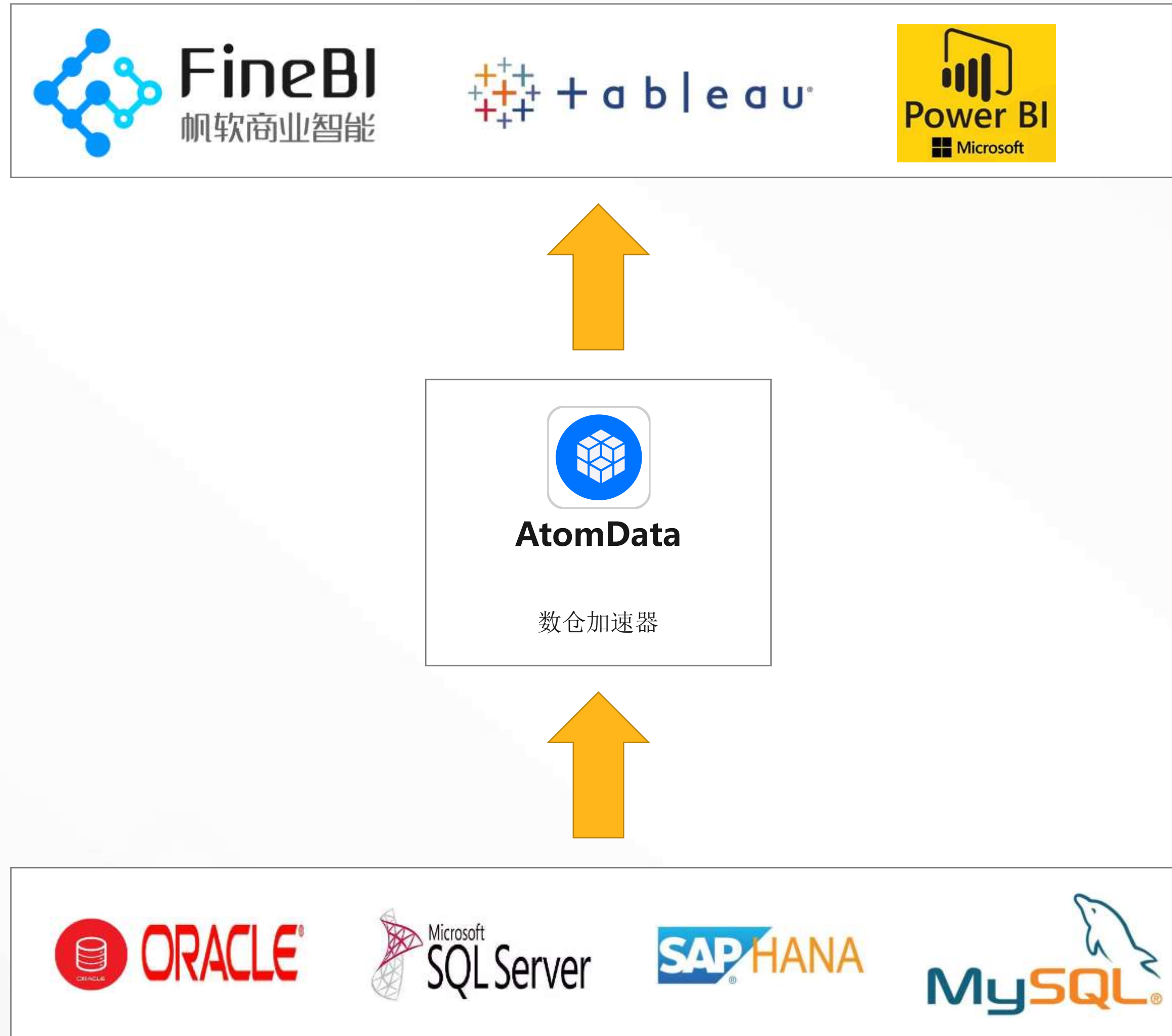
- 支持传统MPP架构和存算分离架构
- 存算分离增强了系统的弹性计算能力

关键技术

丰富的索引和改良的CBO

- 支持丰富的索引类型，可以在不同场景下通过索引提升性能
- 增强的CBO优化器具备存储感知能力
- 提升了即席查询和分析处理的性能

AtomData 最佳实践 — BI加速器



场景问题

随着业务的不断发展，企业决策者们对经营数据有越来越高的要求。既要敏捷的调整报表分析维度，又要从月/周/天的离线报表分析提升到小时/分钟级的实时分析。

往往对于Oracle/SqlServer/SAP HANA/MySQL这些传统数据库来说，做大表复杂分析是一个巨大的性能挑战。延迟往往从十分钟级到小时级，数据库的性能瓶颈成了制约决策效率的关键。

AtomData 解决办法

AtomData是专为数据分析场景设计研发的数据库产品。其如下特性：

- 分布式并行处理架构
- 列式存储引擎
- 向量化计算引擎
- CBO + RBO优化器

利用数据同步工具，将数据实时的从业务端实时同步至AtomData数据库，BI平台直接从AtomData取数。可以实现秒级到亚秒级别的查询延迟，性能提升百倍。

AtomData 最佳实践 — 离在线一体数仓



场景问题

中小企业在建设数据仓库时，往往会按照Hadoop的数仓建设方法论，搭建一套庞大的Hadoop集群并维护众多组件。

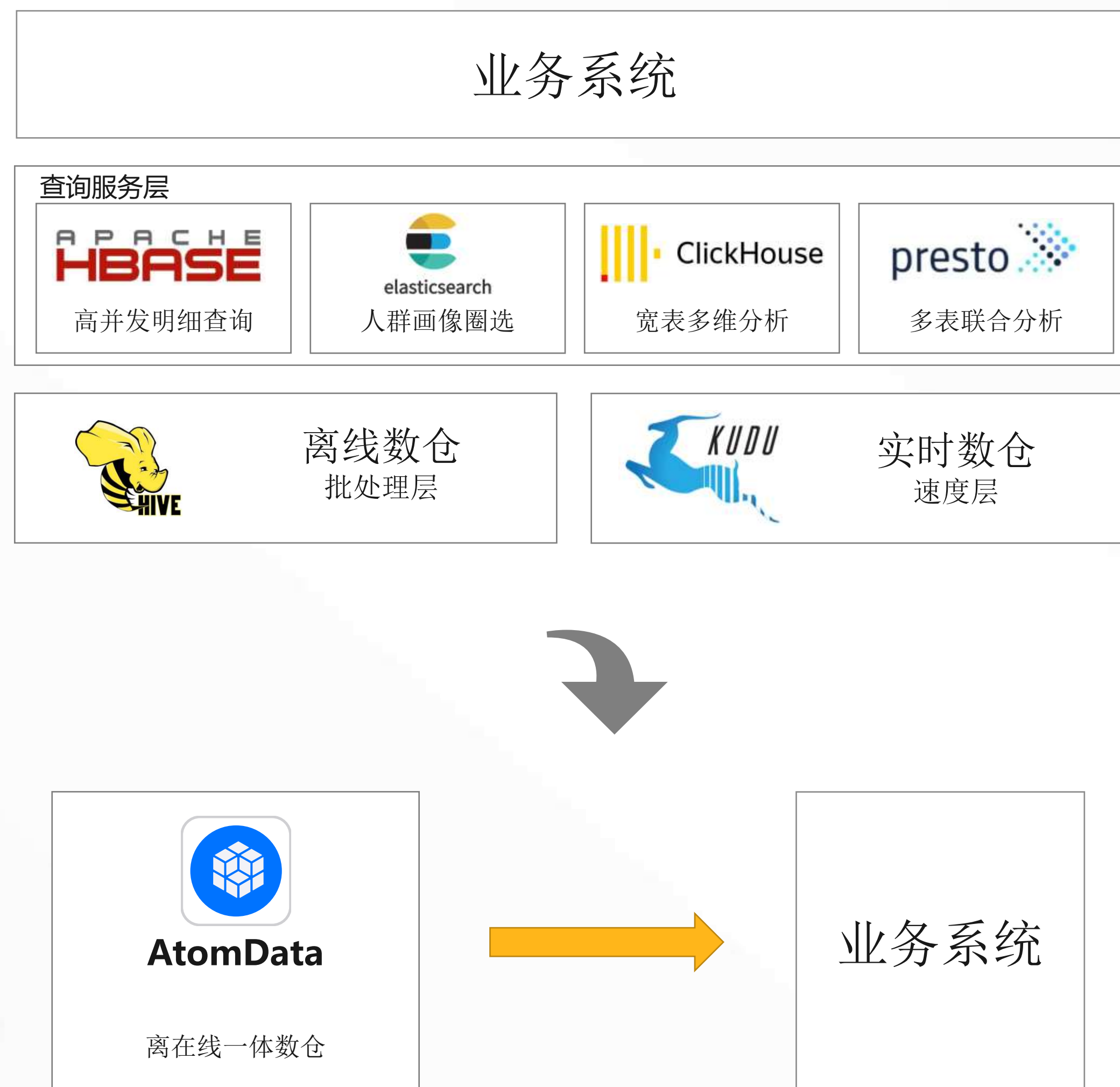
- Hadoop集群庞大，需要投入较高的硬件或租赁成本；
- 技术路线复杂、数据开发链路长，开发周期长；
- 需要较多的开发人员；

这对于中小企业来说投入产出比是不高的。

AtomData 解决的问题

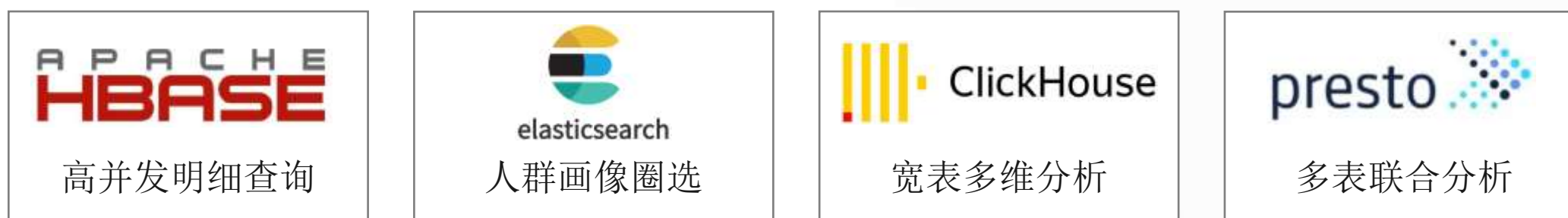
在面对结构化和非结构化数据仓储和分析需求时，AtomData 同时具备离线数仓强大的离线计算能力，高速的实时查询能力。

- 存储计算分离，从容应对存储和计算的弹性扩容需求；
- 资源隔离，让不同的工作负载互不影响；
- 冷热分离，大幅减少冷数据的存储成本；
- MPP计算引擎，满足高并发低延迟的实时查询需求；
- Spill To Disk，计算数据在超出内存后自动溢出到磁盘，让批处理更加稳定；



场景问题

业务系统



离线数据仓库

为了满足固定报表业务的多样的需求，通常会采用多种技术栈融合的手段来完成数据的存储。对于通用的固定报表业务，可以抽象出以下的技术栈解决方案：

- 对于高并发的业务，通常选择 Druid 或者将结果存储在 MySQL 中
- 对于报表类型的业务，通常选择 Impala 或者 Presto
- 对于复杂聚合类操作，通常选择 ClickHouse
- 对于明细查询，通常选择 Elasticsearch

AtomData 为什么快

- 前缀索引可以快速地过滤数据，减少数据扫描量，节省硬件资源地开销
- 向量化引擎可以提供亚秒级查询速度，确保极致性能
- 分区分桶功能可以增加数据指向性，大幅度提升集群并行性
- 物化视图功能可以在明细查询地基础上做不同维度地聚合操作



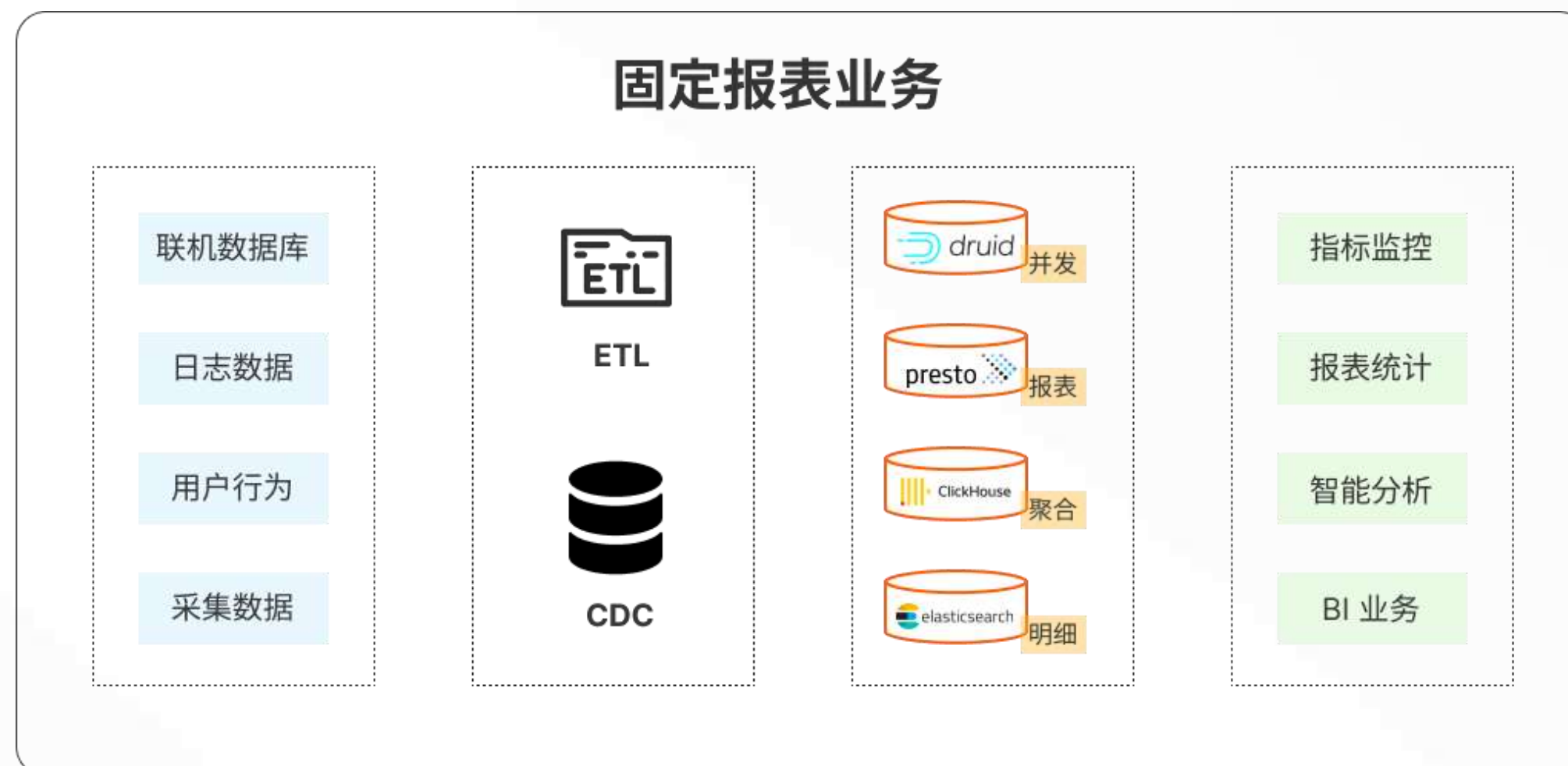
AtomData 最佳实践— 离在线一体

场景问题

为了满足固定报表业务的多样的需求，通常会采用多种技术栈融合的手段来完成数据的存储。对于通用的固定报表业务，可以抽象出以下的技术栈解决方案：

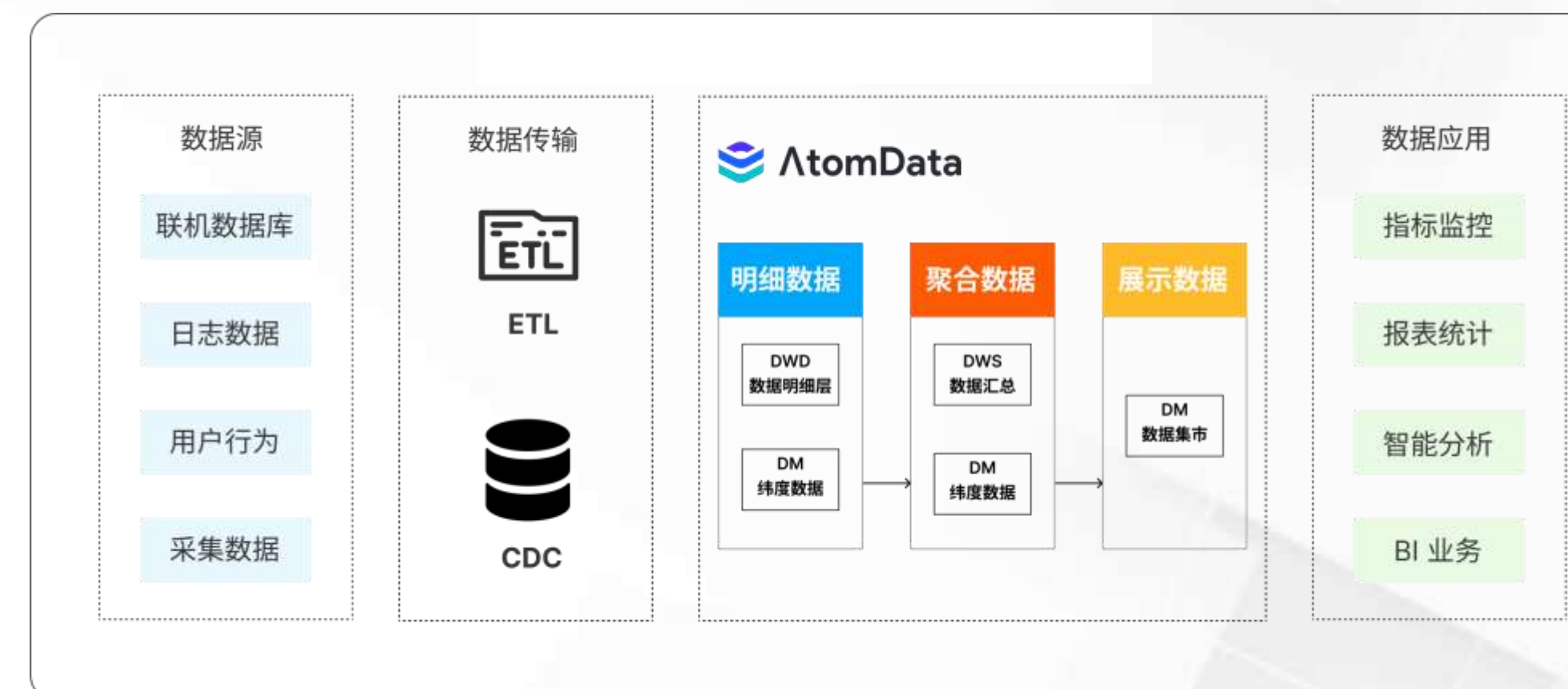
- 对于高并发的业务，通常选择 Druid 或者将结果存储在 MySQL 中
- 对于报表类型的业务，通常选择 Impala 或者 Presto
- 对于复杂聚合类操作，通常选择 ClickHouse
- 对于明细查询，通常选择 Elasticsearch

固定报表业务



AtomData 解决办法

- 前缀索引可以快速地过滤数据，减少数据扫描量，节省硬件资源地开销
- 向量化引擎可以提供亚秒级查询速度，确保极致性能
- 分区分桶功能可以增加数据指向性，大幅度提升集群并行性
- 物化视图功能可以在明细查询地基础上做不同维度地聚合操作



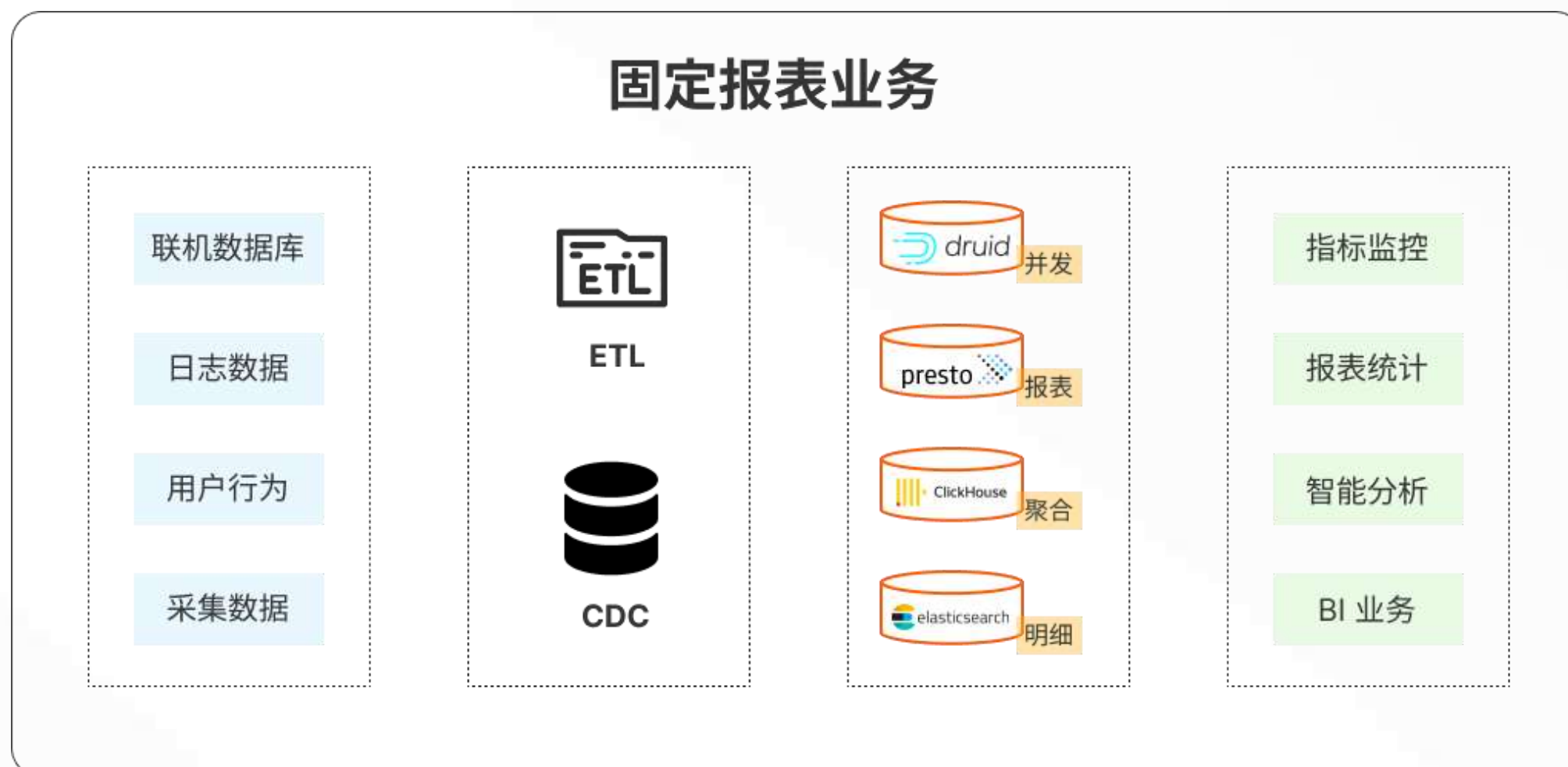
AtomData 典型场景一 报表平台支持

场景问题

为了满足固定报表业务的多样的需求，通常会采用多种技术栈融合的手段来完成数据的存储。对于通用的固定报表业务，可以抽象出以下的技术栈解决方案：

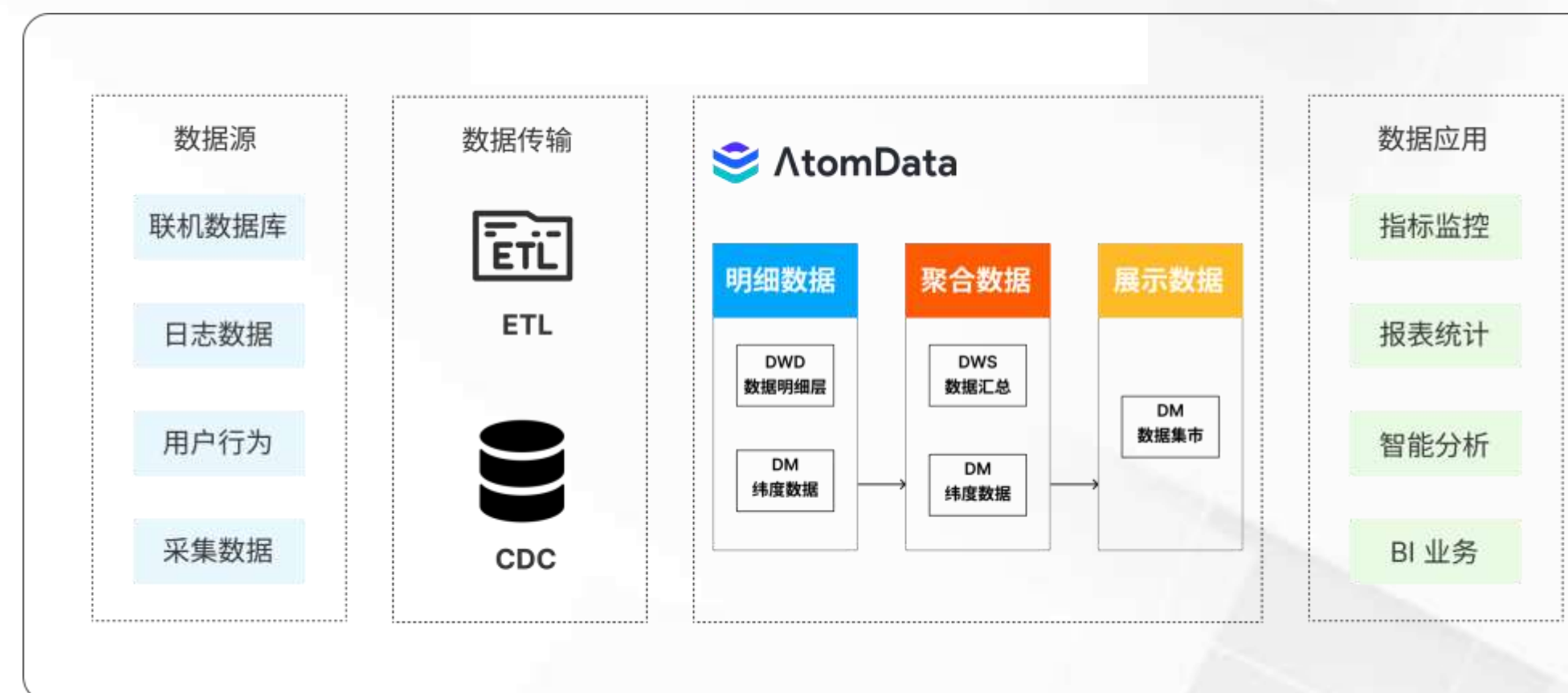
- 对于高并发的业务，通常选择 Druid 或者将结果存储在 MySQL 中
- 对于报表类型的业务，通常选择 Impala 或者 Presto
- 对于复杂聚合类操作，通常选择 ClickHouse
- 对于明细查询，通常选择 Elasticsearch

固定报表业务



AtomData 解决办法

- 前缀索引可以快速地过滤数据，减少数据扫描量，节省硬件资源地开销
- 向量化引擎可以提供亚秒级查询速度，确保极致性能
- 分区分桶功能可以增加数据指向性，大幅度提升集群并行性
- 物化视图功能可以在明细查询地基础上做不同维度地聚合操作

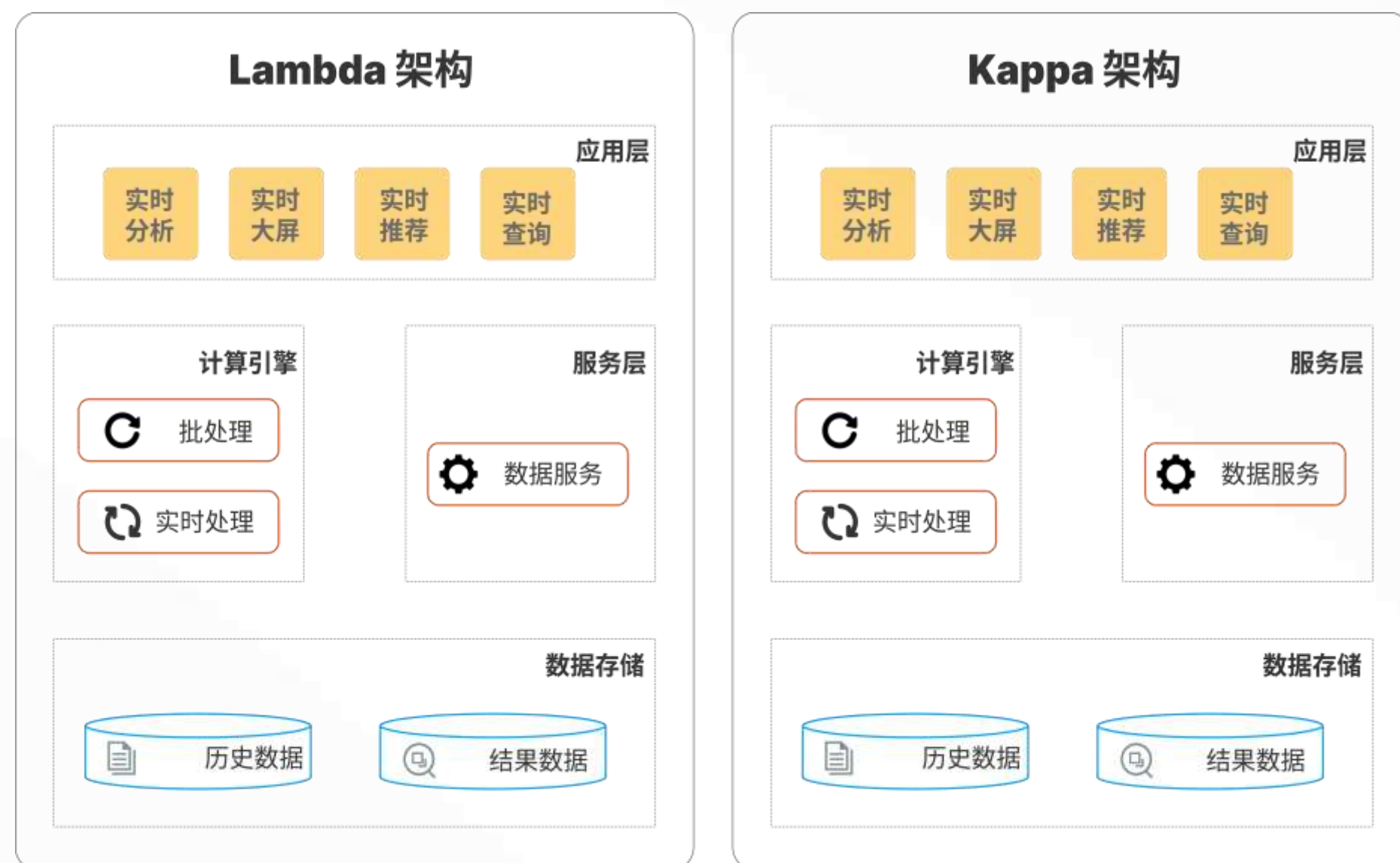


AtomData 典型场景一 实时数仓

场景问题

Lambda 架构在离线数仓的基础上叠加了实时数仓的部分，使用流式引擎处理实时性较高的数据，最后在服务层中统一结果供应用使用。Kappa架构在 Lambda 架构的基础上进行改进，移除了离线数仓的部分，通过重启流式计算实例的方式重新进行全量数据的处理，从而实现离线数仓的计算。

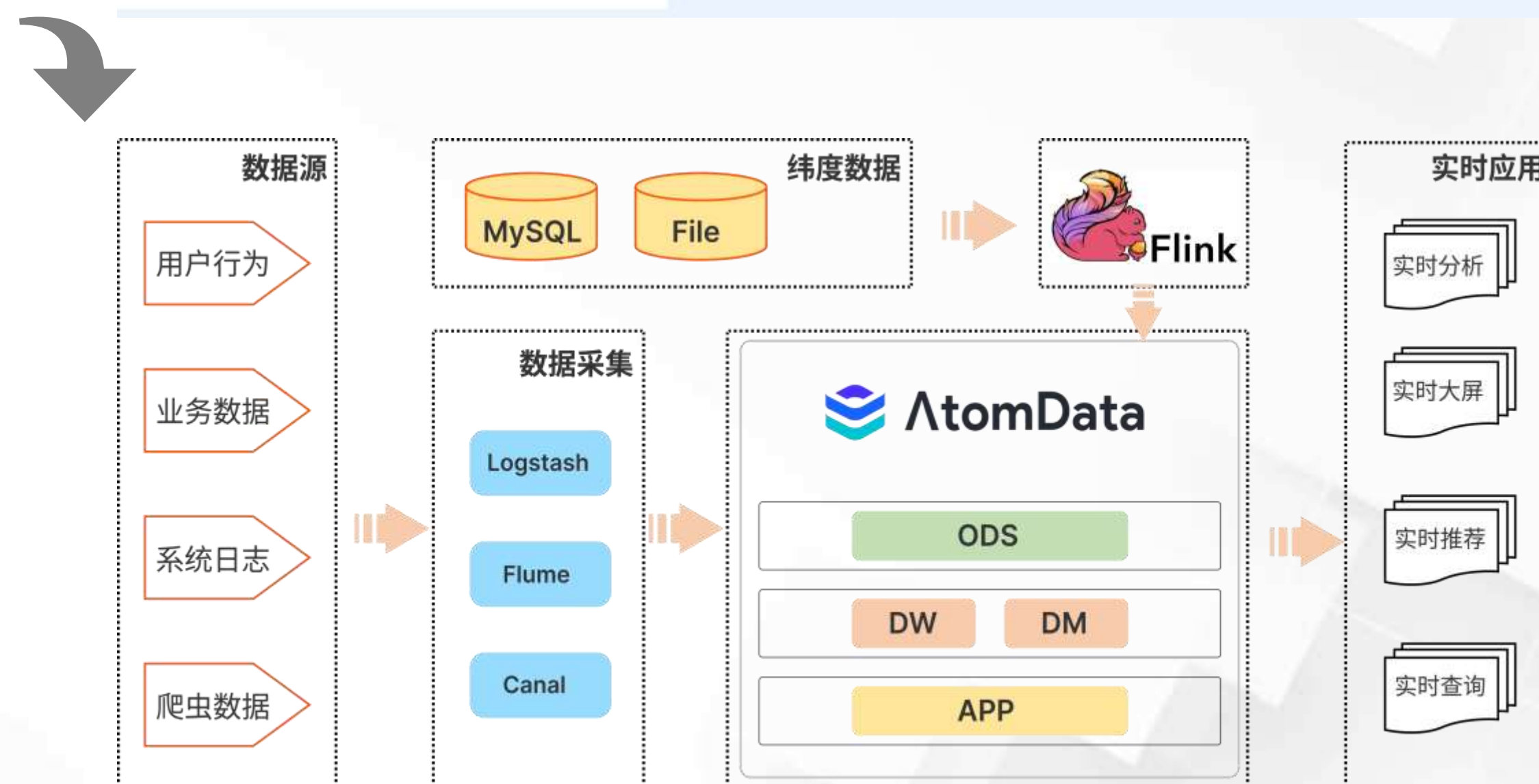
- 消息队列难以保证数据的顺序性，可能出现流式 join 乱序的情况
- 消息队列难以实现 Exactly Once 语义，数据可能会丢失重复
- Lambda 与 Kappa 的架构常会伴随数据孤岛的问题



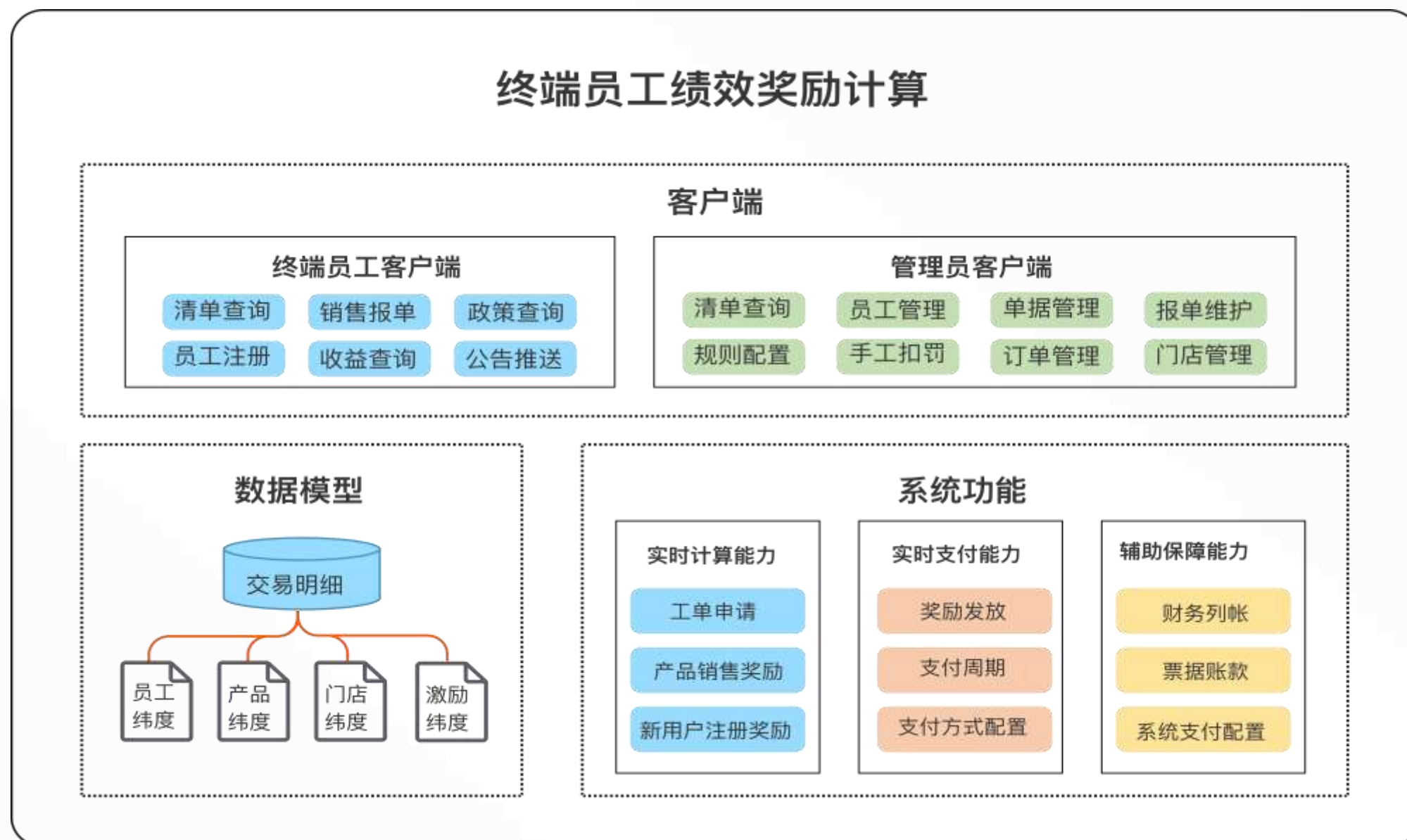
AtomData 解决办法

AtomData 提供了 Flink-Connector 插件，结合 Flink，可以构建高效稳定的实时数仓：

- Flink-Connector 插件提供 Exactly Once 语义
- AtomData 具有多种数据摄入模式，保证导入实时性
- AtomData 提供逻辑视图及物化视图功能，完成数据的 point update 与 delete 操作



终端员工绩效奖励计算

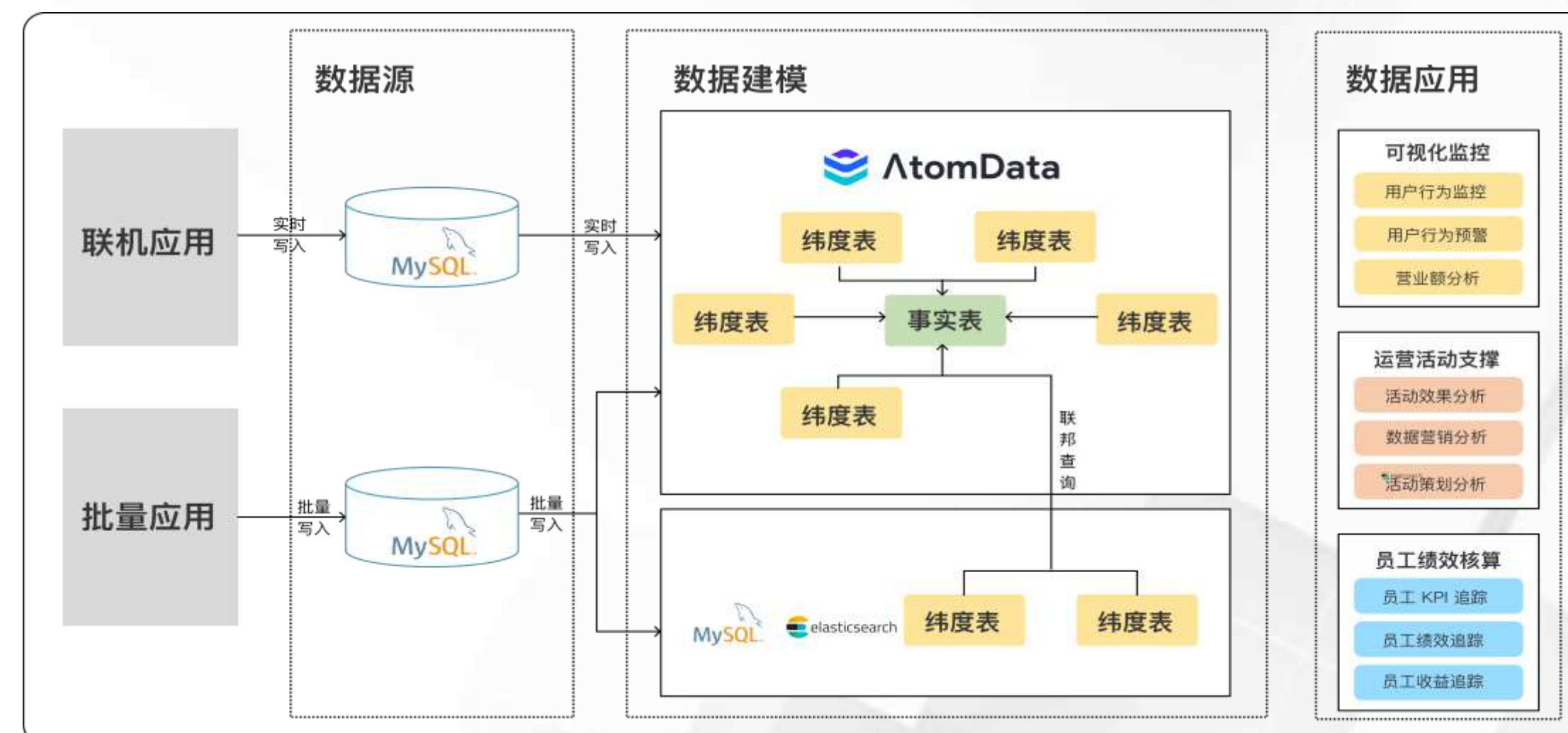


场景问题

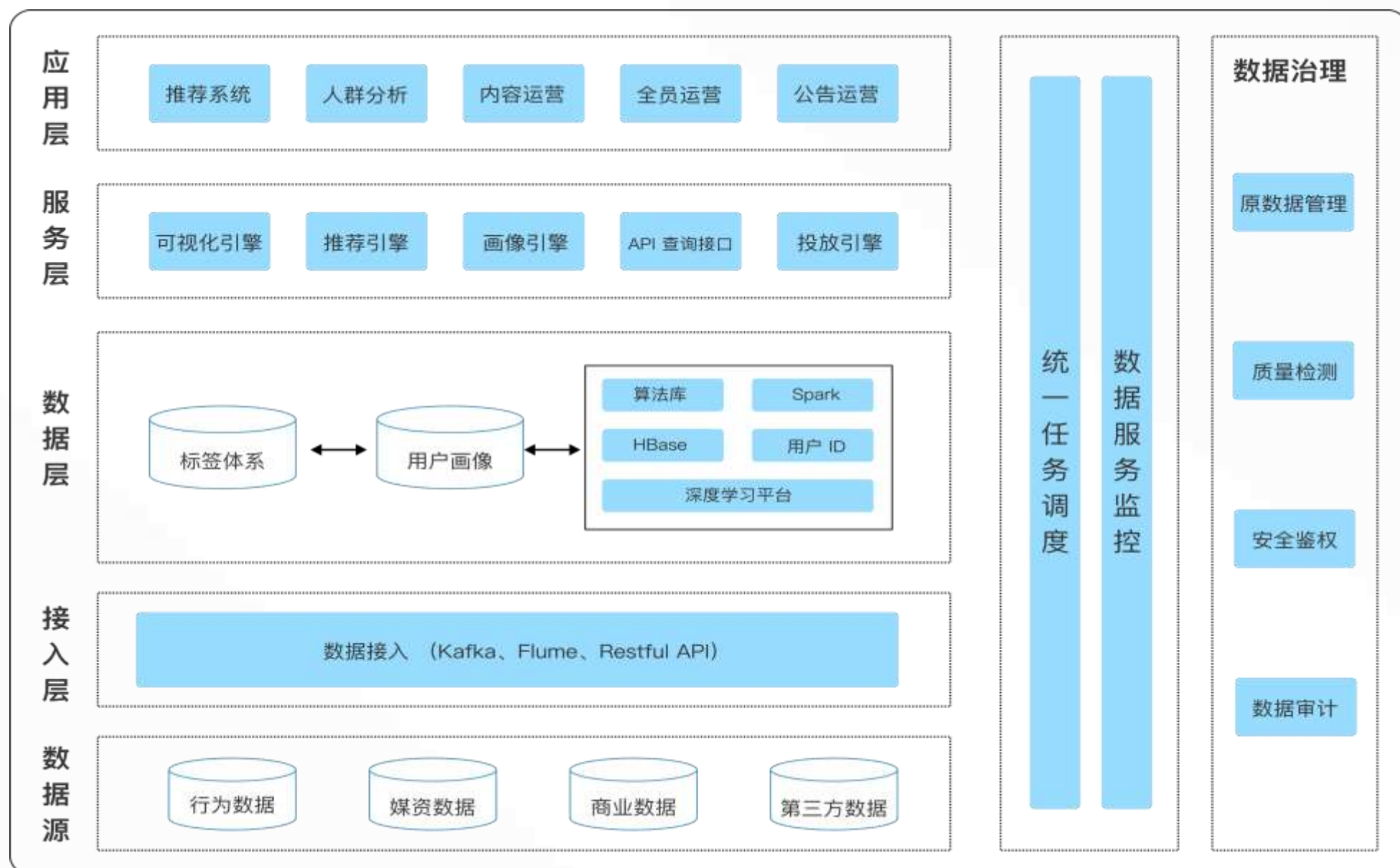
在末端运营业务场景中，一般会根据公司等组织架构分为多级，尤其是对劳动密集型对企业而言，人员流动频繁，对应数仓模型中，就会有大量的纬度变更，传统的大宽表模式或通过 cube 实现预聚合的模式难以应对复杂的业务逻辑，只能通过星型模型分别记录终端员工、报单、支付等过程信息，方便管理员进行多维度等统计分析。奖励计算的及时性不足，审批时间长，报账、兑付环节复杂且人工干预多，会大幅度降低激励的作用。数据导入与模型计算的性能不足，已成为末端运营平台最大的瓶颈。

AtomData 解决办法

- 数据模型灵活，支持多种建模方式来适配复杂业务逻辑
- 现代化物化视图可以完全复杂聚合逻辑，提供极致查询性能
- 多种数据导入方式，确保导入秒级延迟，导入即可用
- 更新模型可以支撑大吞吐量去重，更新等操作，进一步支撑



AtomData 典型场景一 用户画像

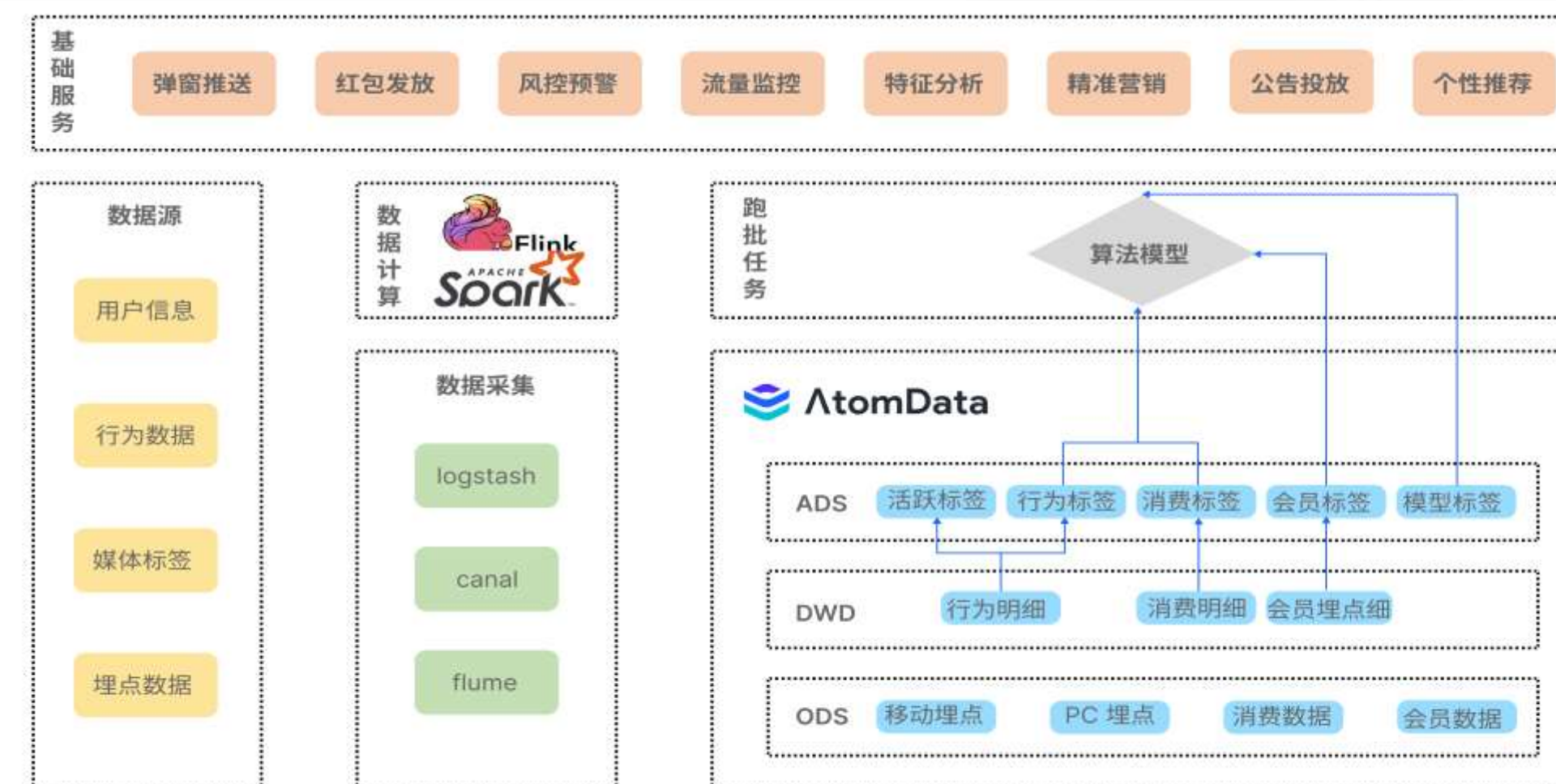


AtomData 解决办法

- 通过 Bitmap 集合计算，可以利用 Bitmap 索引进行精确去重；
- 通过函数支持业务上行列转换需求；
- 通过前缀索引加速业务查询，可以根据用户唯一标识快速定位数据；
- 物化视图功能可以在明细查询地基础上做不同维度地聚合操作

场景问题

- 用户画像平台会从不同的数据源中采集事实与维度数据，然后对这些数据进行深度的、多层次的加工。
- 如何打通多渠道用户信息与多渠道产品信息成了用户画像的难点之一。
- 海量的用户数据对于数据仓库引擎的性能及存储能力提出了极高的要求。
- 在记录和存储亿级画像的过程中，后端数据存储需要能够支撑不断增加的维度和用户偏好，实现秒级入库与更新，支撑个性化推荐，广告投放和精细化运营等业务。



基础服务: 弹窗推送, 红包发放, 风控预警, 流量监控, 特征分析, 精准营销, 公告投放, 个性推荐

广泛、深入的生态融合，数据智能驱动创新



(1)对象存储: Minio、S3、OSS 

(2)数据库管理工具: Navicat Premium、Datagrid、DBVisualizer、Dbeaver、SQL WorkBench、mysql client



(3)BI工具: FineBI、永洪BI、Tableau、Grafana、Superset、有数BI、DataV、QuickBI、QlikView、FineReport、PowerBI、SmartBI、Sugar BI



(4)数据同步: Datax、canal、Kettle、mysql-client8等 

(5)数据组件: hadoop生态圈、Kafka、Flink、Spark、LogStash、ADB组件



(6)调度工具: Dolphin scheduler、DataWorks、Airflow、Azkaban、Oozie、Activiti、Easyscheduler、Flowable



(7) 数据备份: Mysqldump



(8)数据库对接: Oracle、DB2、SQL Server、MySQL、PolarDB、MQ、MariaDB、Druid、Presto、Hive、Kafka、Redis、MongoDB、ClickHouse、Trino



(9)三方编程开发支持: Java、NodeJS、Python、Golang



(10)外表关联: Mysql、Hive、ElasticSearch



(11)地图厂商适配: 四维地图、京东地图



(12)数据可视化: DataEase



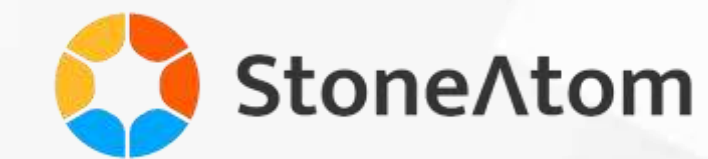
(13)数据中台对接: 数澜、华傲、袋鼠云、半云



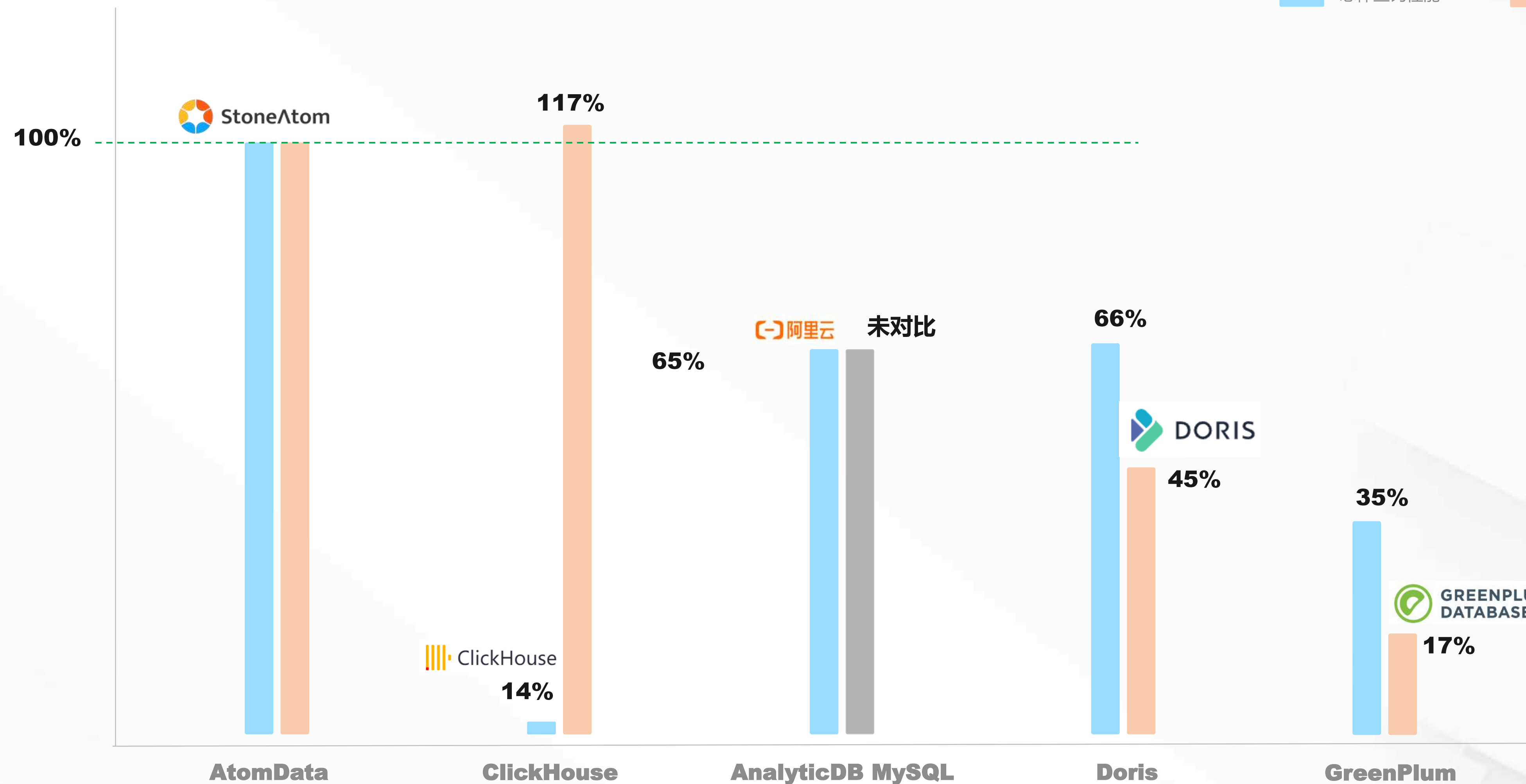
(14)大屏展示对接: 四方伟业等



AtomData 性能对比



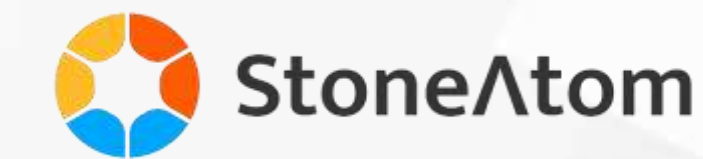
总体查询性能 单表查询性能



测试方法:

- 使用TPC-H/TPC-DS 的脚本测试，依次执行TPC-H/TPC-DS测试集。
- 测试前，均执行两轮测试集作为预热环境。连续执行三次测试，取第三次测试结果的值；

行业案例：某企业用工招聘平台核心业务数据平台



该企业创立于2013年，是国内领先的一站式灵活用工招聘平台。为65万+企业提供兼职招聘与管理服务，有效降低用工成本，提升经营效率。已为超3800万年轻人提供4.5亿+次兼职机会，单日报名人数超85万，服务范围覆盖全国300多个城市，位居行业第一。

痛点和需求

充足的计算与存储能力

传统商业数据仓库扩展性差，难以应对业务应用对容量和算力日益增长的需求，数据平台需要以合理的成本提供充足的资源供给。

高质高效的数据应用环境

改善数据分散和烟囱式建设现状，以解决实际业务问题为出发点，建设企业级的数据能力复用平台，沉淀可复用易扩展的数据资产。

自主可控赋能业务创新

建设安全可靠、技术领先、自主可控的数据平台，深度挖掘数据资产的价值，实现数据驱动经营。

该企业采用石原子一站式企业级实时数仓AtomData替代了以前采用的传统商业型数据库加云厂商商业服务，承载了平台上所有数据应用，提升数据分析效能，赋能业务即时决策。从根本上解决了业务数据不一致，实时业务不稳定的问题；简化数据平台架构，非专业人员轻松用数，在业务高峰时轻松的扩展计算节点应对突发的业务大流量挑战。

核心业务数据平台



云原生管控	数据处理	数据管理
鉴权访问	离线计算	规则引擎
计费服务	实时计算	数据质量
负载均衡	查询分析	数据安全
资源管控	融合分析	智能优化

一体化架构设计

10+ 组件 → 1 平台

更大存储能力

TB级 → 10PB

更快分析速度

316% ↓

数据分析时长

更低成本

48% ↓

数据存储计算成本

行业案例：助力某区政府实现“数字服务”一体化能力



该区为中心城区之一，拥有两个城市副中心。根据第七次人口普查数据，该区常住人口119.76万人。2021年，全区GDP总量880.98亿元。

基于石原子“云中立 × 一站式”理念和架构，打造按需供给、据实结算的数据服务模式

- ▶ 以“全区数据统一”为建设理念，打造多云协同、统一管理、集中调度。
- ▶ 采用融合专有云、公有云的混合架构，将对公众服务的创新应用部署到公有云，关键数据存储于专有云上，实现全区生态环保、各专题的一屏统揽、企业信息、一网调度、经济能力一张图。
- ▶ 在基础资源池之上构建一站式数据分析架构，告别了瀑布式开发、龟速上线，依托于StoneDB-X的弹性能力，可以轻松应对“类互联网业务”的峰值。

数据资源中心



统一数据存储

实时查询

融合分析

痛点和需求

业务多，数据量大，政务平台冲击大



按照“十四五”数字经济发展规划，政府要通过加强数字基础设施建设，完善数字经济治理体系，进一步支撑政务服务、公共服务、社会治理和民生保障，逐渐形成政府主导、多元参与、法制健全、安全保障的数据共享生态圈。

随着越来越多的政务数据、流程打通，激活了更多的政务应用创新，越来越多的应用给政务平台带来了流量冲击。传统的政务架构和方式“力有不逮”。



采用云原生的应用开发部署模式，使得业务上线周期从月级下降到**周级**，**3周**完成上线运行，期间迭代超过5个版本，保障了所有功能正常上线。



采用了与公有云联动模式作支撑，实现南岸区数据全生命周期、全业务过程、全价值链管控，数据分析时间从**分钟级降到秒级内**，实现从“电子服务”到“数字服务”的转型。



统一的数据中心，采用StoneDB-X将当前各业务系统数据实现全方位归集处理，实现应用与数据分离、实现**300TB**数据计算存储成本**降低72%**。

更多信息



www.stoneatom.com



service@stoneatom.com



0571-85220891



杭州市余杭区数智引擎创新园

石原子产品，服务于您、服务于您的客户！



StoneDB



AtomData