分布式数据库海量数据存储和实时查询实现与应用

巨杉数据库 乔国治

SequoiaDB
巨杉数据库

公司简介：

SequoiaDB巨杉数据库（广州巨杉软件），成立于2011年专注于新一代企业大数据平台研发，其核心产品SequoiaDB（巨杉数据库）是国内第一款新一代分布式数据库;

核心产品完全自主研发，数据库引擎没有基于任何开源数据库源代码，已经成功部署并运行在多家世界500强企业的生产环境中;

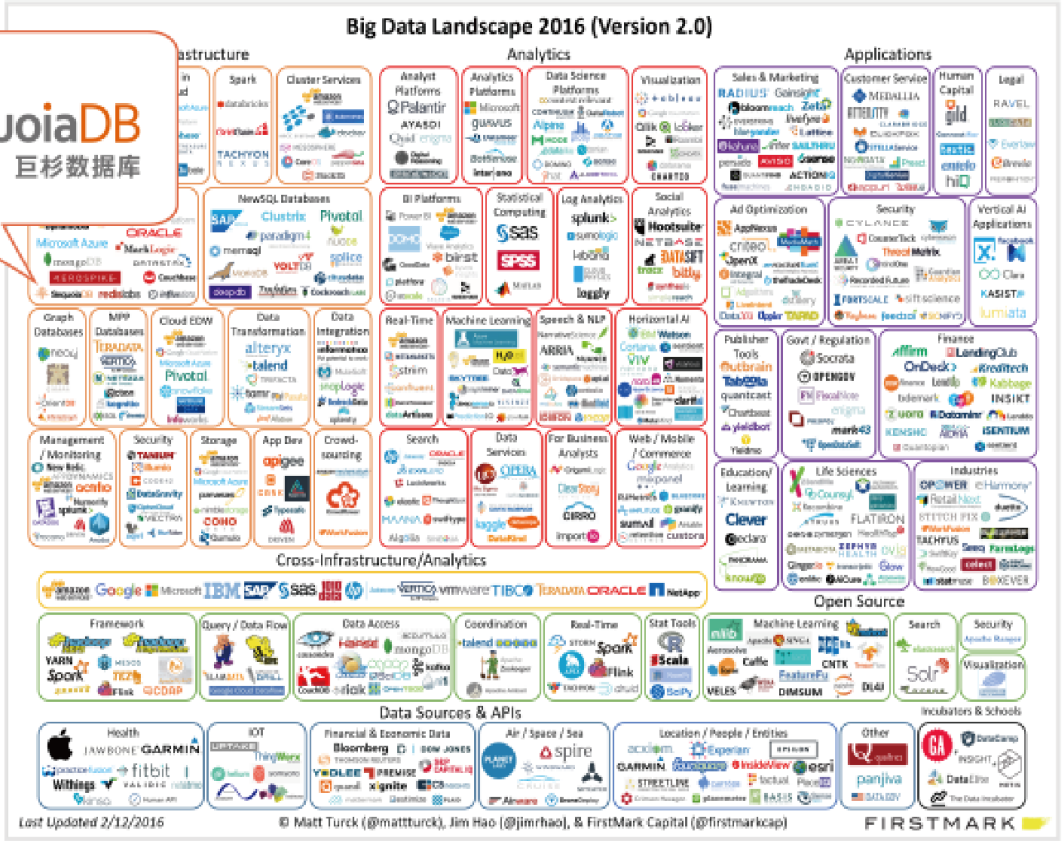
获著名基金启明创投（A轮）与DCM（B轮）融资;

中国第一款商业开源数据库产品
www.github.com/sequoiadb/sequoiadb
www.sequoiadb.com
www.oschina.net/p/sequoiadb

"全球创新企业Top100"
——《红鲱鱼》
美国最具影响力商业媒体

"中国创新企业50强"
——《快公司》
美国著名创新媒体

SequoiaDB
巨杉数据库



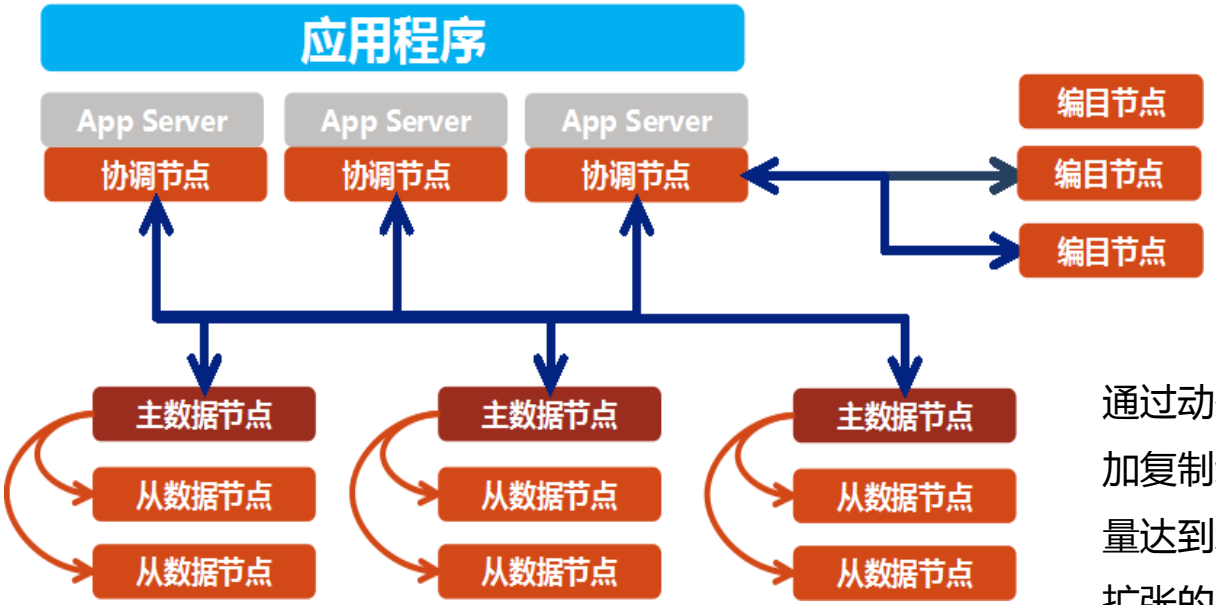
成熟的商业化自主研发数据库 — 行业用户认可

- 主要客户以金融、运营商、政府、交通航空、互联网等行业为基础
- 研发中心在深圳，现场支持队伍部署在北京、上海、广州三地

海量数据存储

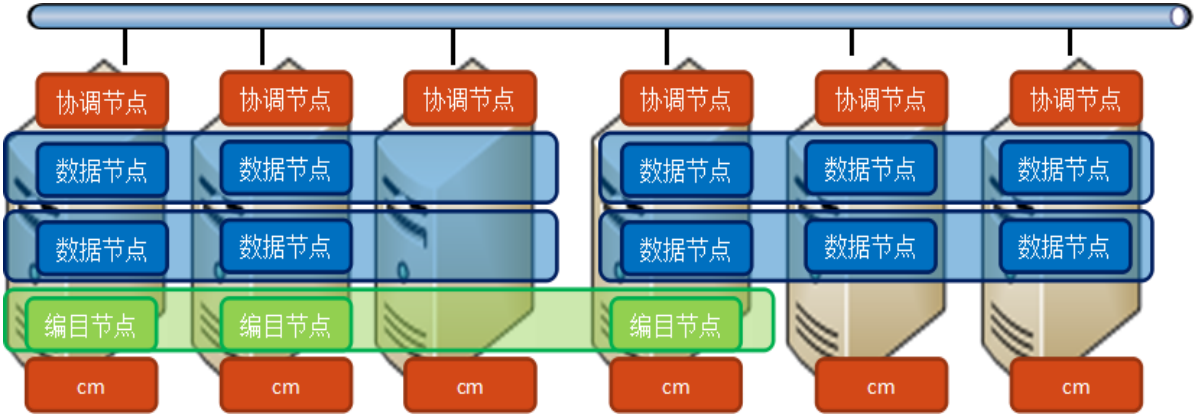
SequoiaDB

SequoiaDB分布式数据库架构



通过动态增加复制组数量达到水平扩张的目的

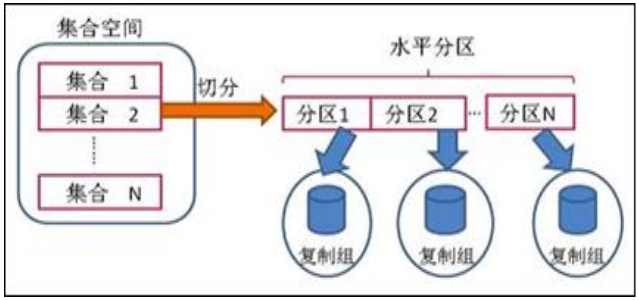
SequoiaDB物理架构

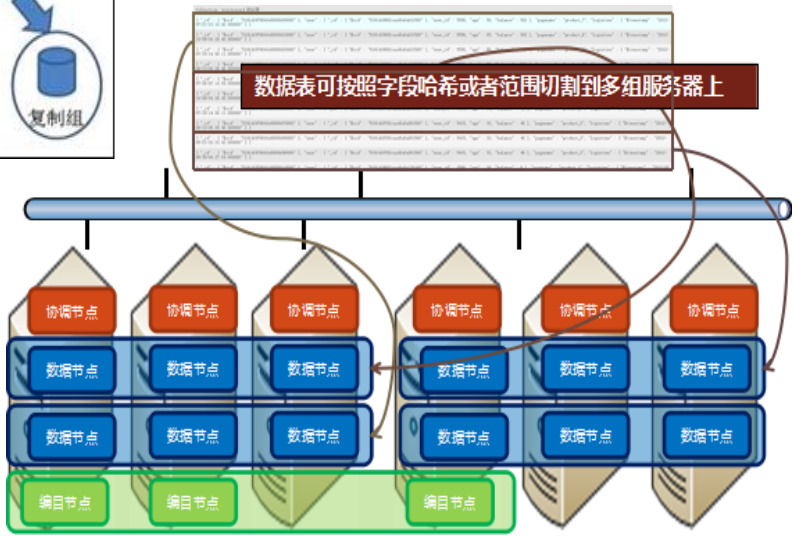


角色	功能
协调节点	胖客户层，从编目读取数据分布信息，从数据节点读取数据
编目节点	负责元数据信息存储，包括组信息、表切割信息
数据节点	负责数据表存储，提供查询、聚集、数据复制功能
CM节点	负责集群管理，包括watchdog, 节点增删启停

数据水平分区

SequoiaDB支持水平分区和垂直分区。水平分区尽可能选择唯一性较高的字段





优势： 容量和性能可线性扩展

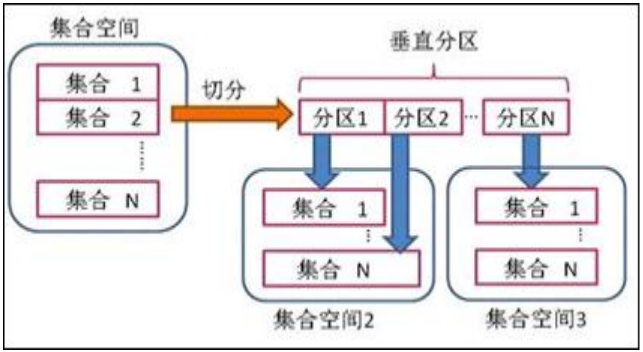
数据垂直分区

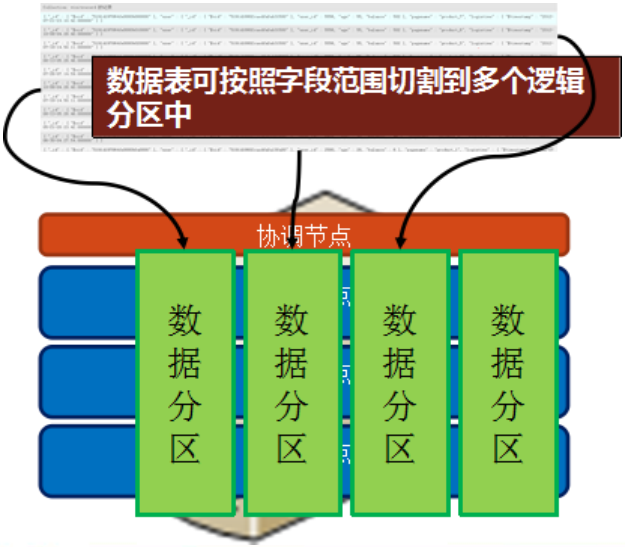
垂直分区选择记录生成的时间戳作为分区字段

-保障最新的一段时间数据尽可能驻留内存

-可以按照时间范围快速删除数据

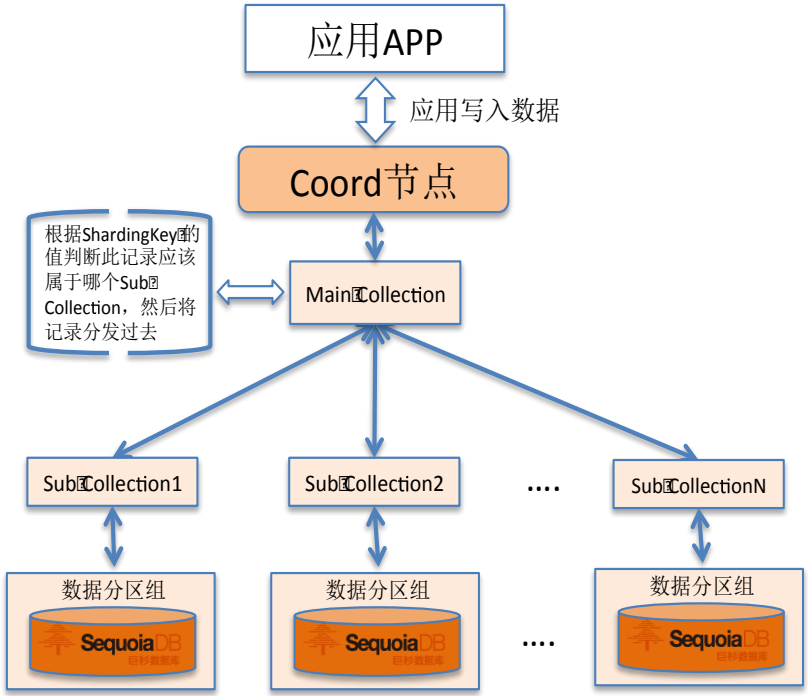
优势： 容量和性能可线性扩展





举例：Partition 分区机制

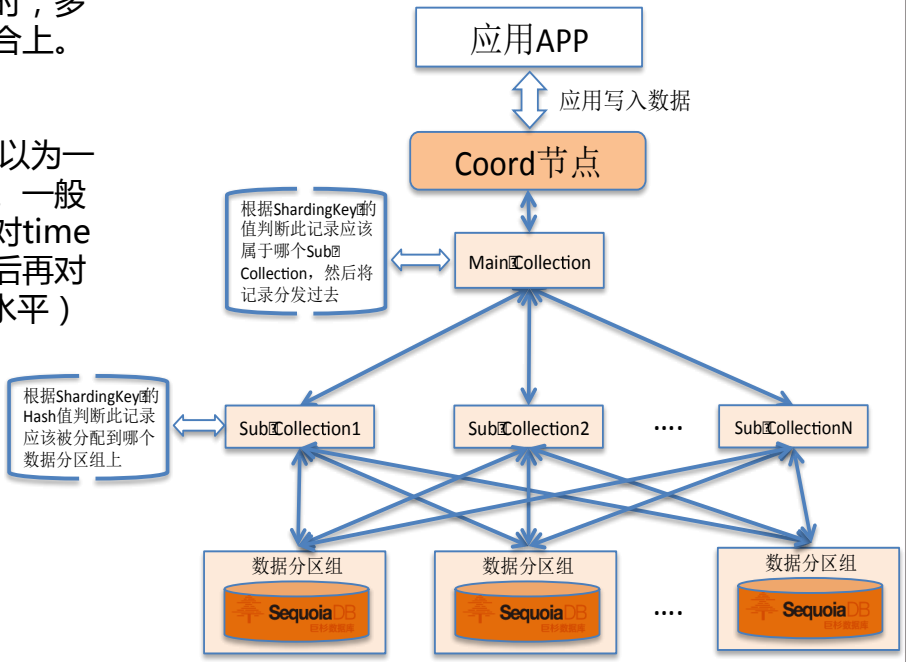
这个功能与部分关系型数据库的Partition功能类似，都是在数据库中建立一张逻辑的总视图，然后将多个Partition通过某个字段的范围限定挂载到总视图上。main collection为逻辑视图，只在编目节点中保留一些数据范围信息，而sub collection 则是数据库中普通建立的集合，只是在配合main collection 一起使用时，才被称呼为sub collection。



举例：多维分区机制

在SequoiaDB数据库多种分布式原理上，最为复杂和最为高效的方法，就是数据多维分区。多维分区，顾名思义，就是在对集合做数据分区时，多种分区方式同时作用在一个集合上。

如图所示
目前SequoiaDB的多维分区可以为一个集合同时提供两种分区方式。一般情况下，用户可以在主子表中对time字段做（垂直）范围切分，然后再对sub collection 的id字段做（水平）Hash切分。



加速实时查询

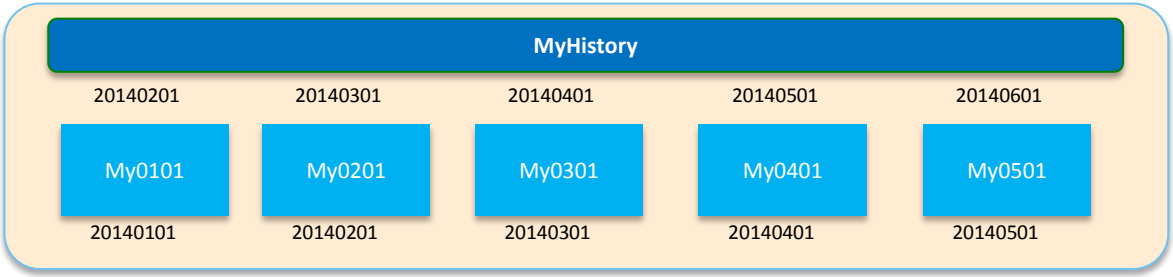
主子集合/时间序列机制

- 时间序列数据的处理是常见需求，很多应用中海量数据的时间特性以递增为主，旧数据的热度随着时间的推移递减
- SequoiaDB提供集合分区（主子集合）机制可以轻松应对时间序数据
  - 避免单一集合数据量膨胀时索引树过大而导致的写入性能雪崩
  - 按时间序能直观反映数据访问热点，保障热点数据集合的性能
  - 直观的分配资源给不同集合，直观的备份、归档规则

```
db.cs.createCL( "My0101" )
db.cs.MyHistory.attachCL(
    "cs.My0101",
  {UpBound:{date:" 20140201" },
   LowBound:{date:" 20140101" }}
)
```

```
db.cs.createCL( "My0201" )
db.cs.MyHistory.attachCL(
    "cs.My0201",
  {UpBound:{date:" 20140301" },
   LowBound:{date:" 20140201" }}
)
```

```
db.cs.createCL( "My0301" )
db.cs.MyHistory.attachCL(
    "cs.My0301",
  {UpBound:{date:" 20140401" },
   LowBound:{date:" 20140301" }}
)
```



高性能索引

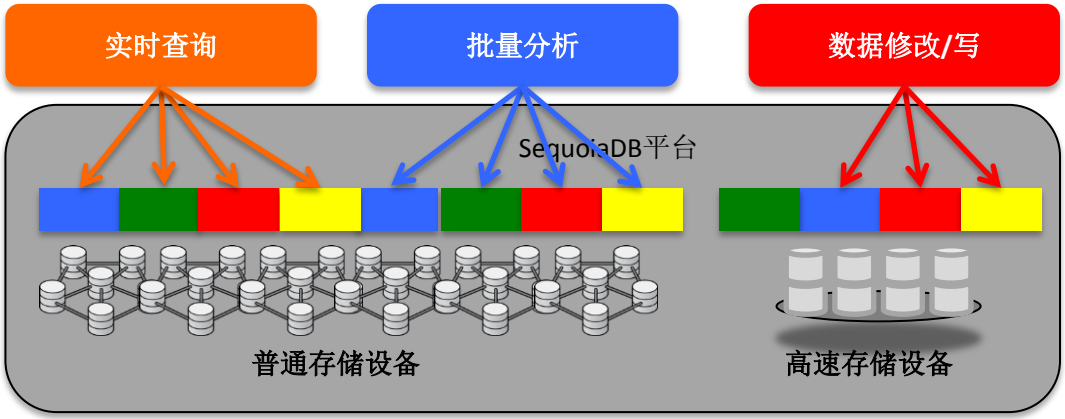
在SequoiaDB中，既支持单字段索引也支持联合索引（多字段索引），可以适配复杂的查询条件。

每个集合支持多达64个索引， 可根据应用的需求灵活定义多个索引，同时为多种查询提速。

索引数据可指定创建在单独的高性能存储上（比如SSD），进一步提高查询效率。

读写分离及自定制数据分布策略

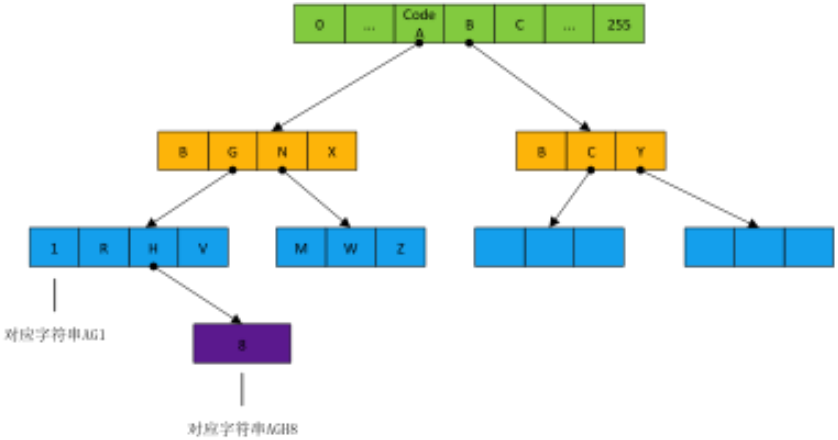
- 数据在多个分布节点内自动复制，并实现写请求和读请求的自动分离，避免读请求对数据写入的影响。
- 此外，可进一步定制数据分布策略，保证不同类型业务可以运行在同一平台上，但同时又不会互相干扰，比如：
  - 冷/热数据区分离
  - 写交易的“强一致性”和“弱一致性”分离
  - 查询/批量分离



高效压缩机制

SequoiaDB支持Snappy和LZW两种压缩机制，既能实现快速压缩，也能满足深度压缩需求。

在IO吞吐量非常高的查询场景下，基于数据字典的深度压缩机制能够大幅降低IO开销，有效提高查询效率。

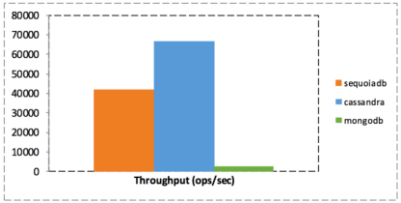


SequoiaSQL引擎特点

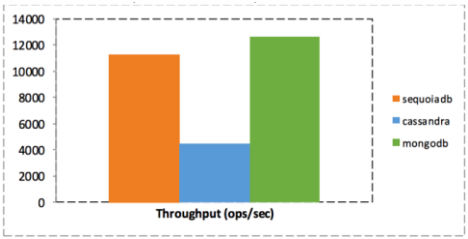
- 兼容标准SQL2003
  - Window Function
  - WITH-AS clause
  - Sub-query
  - Semi-join, Anti-join, Left-join, Right-join, Full-outer-join
  - UDF
- 双引擎
  - 对于OLTP与OLAP使用不同的执行引擎
  - OLTP可以支持高并发实时查询，最高可达十万级OPS
  - OLAP可以支持海量数据分析与交互式报表
- 高性能
  - 新一代大数据SQL执行引擎，使用原生SQL优化执行方式
  - 50倍Hive性能
  - 2-5倍Impala/Spark性能
- 开放性
  - 丰富的扩展性和语言支持(Java , Perl, Python, R, C, etc)
  - 丰富的第三方工具支持(GoldenGate, DataStage, Pentaho Kettle, etc)
  - 丰富的BI工具支持 Tableau, Pentaho, BO, BIEE, Cognos, etc.

第三方性能对比（对比MongoDB、Cassandra）

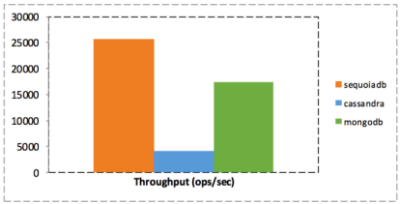
巨杉数据库的技术在业界领先，如今的SequoiaDB 2.6版本更是在各项企业级功能上超越了硅谷同类产品。同时，对比众多硅谷的同类产品，SequoiaDB巨杉数据库在各项性能指标都保持绝对领先。



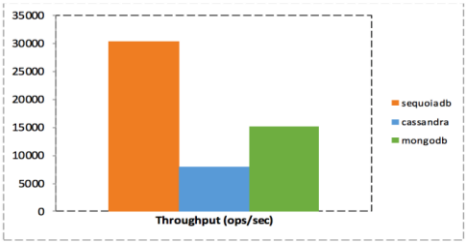
100%写入



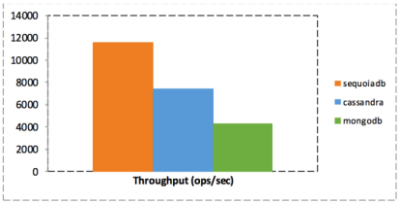
95%读
5%更新



100%读



95%读
5%写入



50%读
50%更新

http://www.bankmark.de/wp-content/uploads/2014/12/bankmark-20141201-WP-[illegible]hark.pdf

应用案例

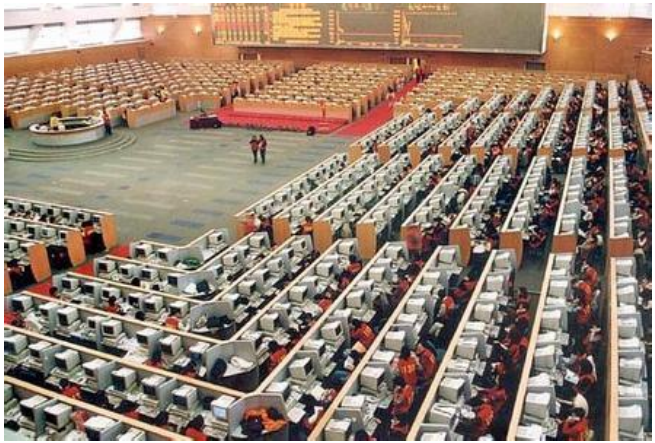
SequoiaDB

证券行业高并发查询

某证券监管机构的股票交易信息管理系统，存储全国交易所每天上传的所有的股票交易信息。如今通过APP，网页端等，开放给股民用户进行实时的查询。

通过搭建基于SequoiaDB的数据库存储，该机构将所有历史数据实现在线化，同时保证每天增量的及时写入。

- 峰值超过2亿条记录写入
- 高峰时段，同时有超过百亿级别的数据需要被检索、调用
- 峰值并发量超过10000
- 高峰时段，查询返回时间小于100ms
- 实际测试性能10倍于原有MySQL
- 操作涉及3张数据表的关联，总量超过1000亿条数据



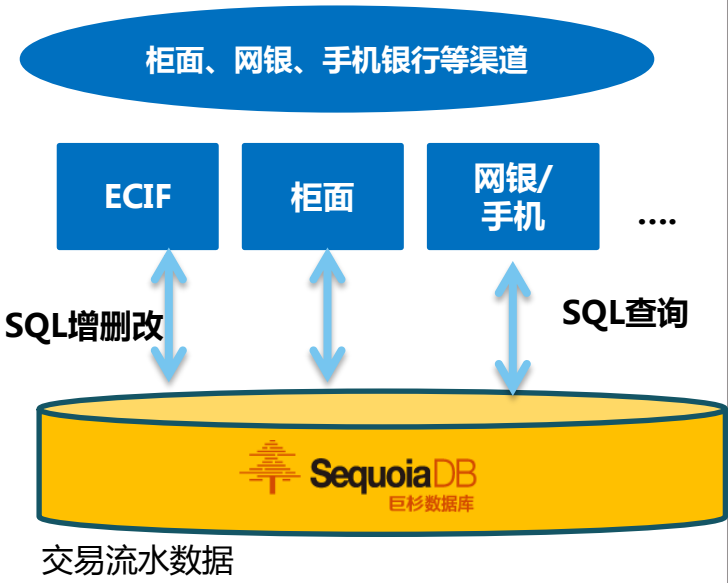
银行历史数据平台

需求

- 高扩展性和稳定性
- 数据分析的接口
- 平滑过渡，不影响原有数据处理流程， 对现有应用影响尽量

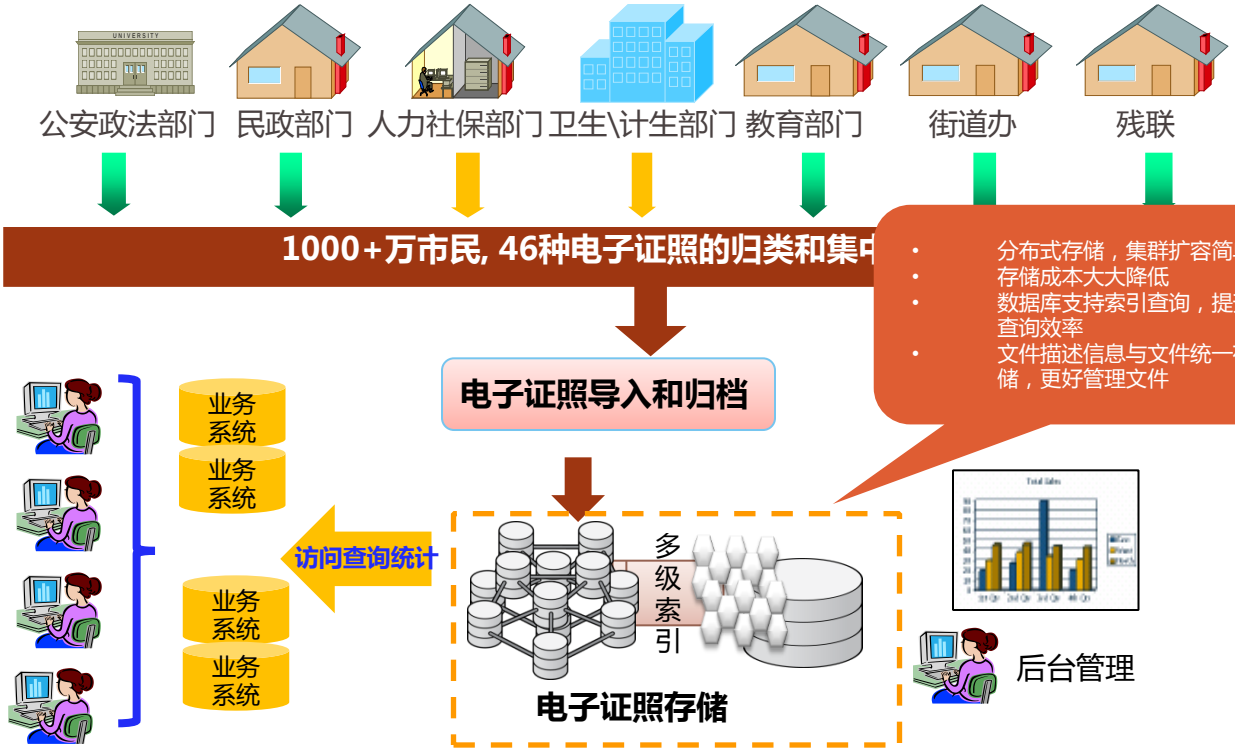
成果

- 超过1PB数据存储，100+节点
- 50亿记录的实时查询性能<1s
- T+1 批量将生产数据全量及增量导入SDB
- SDB对外提供SQL接口，可继续使用现有查询应用



政务数据大数据湖

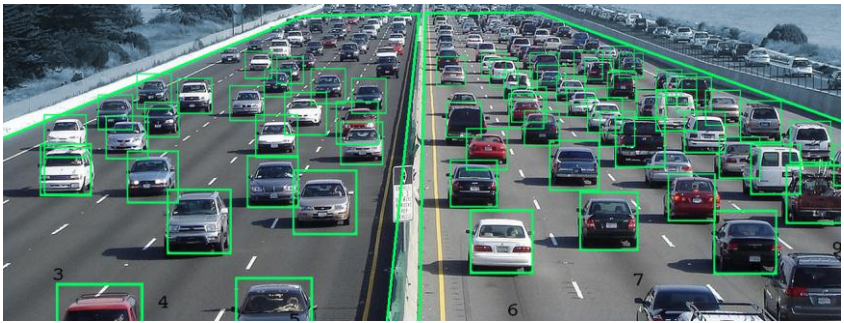
- 一卡通项目中电子证照数据的共享



交通/安防 监控视频影像管理

在公安与交通行业，针对视频卡口的大数据存储、分析与应用一直以来是最受关注的主题。借助SequoiaDB半结构化对象存储、分布式横向扩展能力以及非结构化影像存储引擎，交通部门可以从卡口视频文件中提取出的车牌信息、位置信息、以及时间信息按照三个维度汇总，进行道路拥堵预测、车辆轨迹跟踪、套牌车监控、尾随车辆监控等多种安防措施。

- 兼容各采集系统的数据，统一汇总统一管理，支持多索引多维度查询。
- 并发处理能力高，支持多警种多业务多用户同时使用，系统反应快速。
- 系统架构简单，易于部署和维护，易于横向扩展。
- 存储层和分析层松耦合，均可弹性扩充；数据加载快，分布式计算分析效率高。
- 可灵活配套多种分析工具；SQL通用性高，适合警务操作人员灵活查询；



OTA旅游 /电商 - 多类型数据混合存储

互联网应用的特点，带来了几项重要的挑战。

- 数据量大
- 业务增长快
- 数据类型多样

途牛旅游网“资源系统”的另一个核心业务模块，负责存储和记录所有的旅游方案相关的资源信息，包括酒店，机票，门票，火车票，汽车票，地接，当地服务等。通过使用巨杉数据库，在满足海量存储的同时，也能实现高效的在线资源查询。

SequoiaDB 巨杉数据库
www.sequoiadb.com
Sales_support@sequoiadb.com
hr@sequoiadb.com
400-8038-339

招聘技术大牛

北京/上海/广州/深圳
数据库研发/大数据开发/售前技术支持

大家可以扫描二维码填写有关我们演讲内容的反馈/提问问卷
提交后可以在场外展位领取专属礼品！