

NewSQL数据库技术与企业级应用

巨杉数据库 王涛





SequoiaDB
巨杉数据库

公司简介：

SequoiaDB巨杉数据库，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



“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

Big Data Landscape 2016 (Version 2.0)



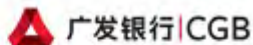
Last Updated 2/12/2016

© Matt Turck (@mattturck), Jin Hao (@jinhao), & FirstMark Capital (@firstmarkcap)

FIRSTMARK

行业与企业级用户

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



大数据数据湖

大数据思维 – 先有数据，再探索关系

大数据的应用特点：**全量、关联、实时**

数据成为真正核心全息资产，巨大的信息量足以支撑以前难以完成的任务

特定场景下基于样本数据的挖掘



以问题为中心
以全量数据为基础的统计



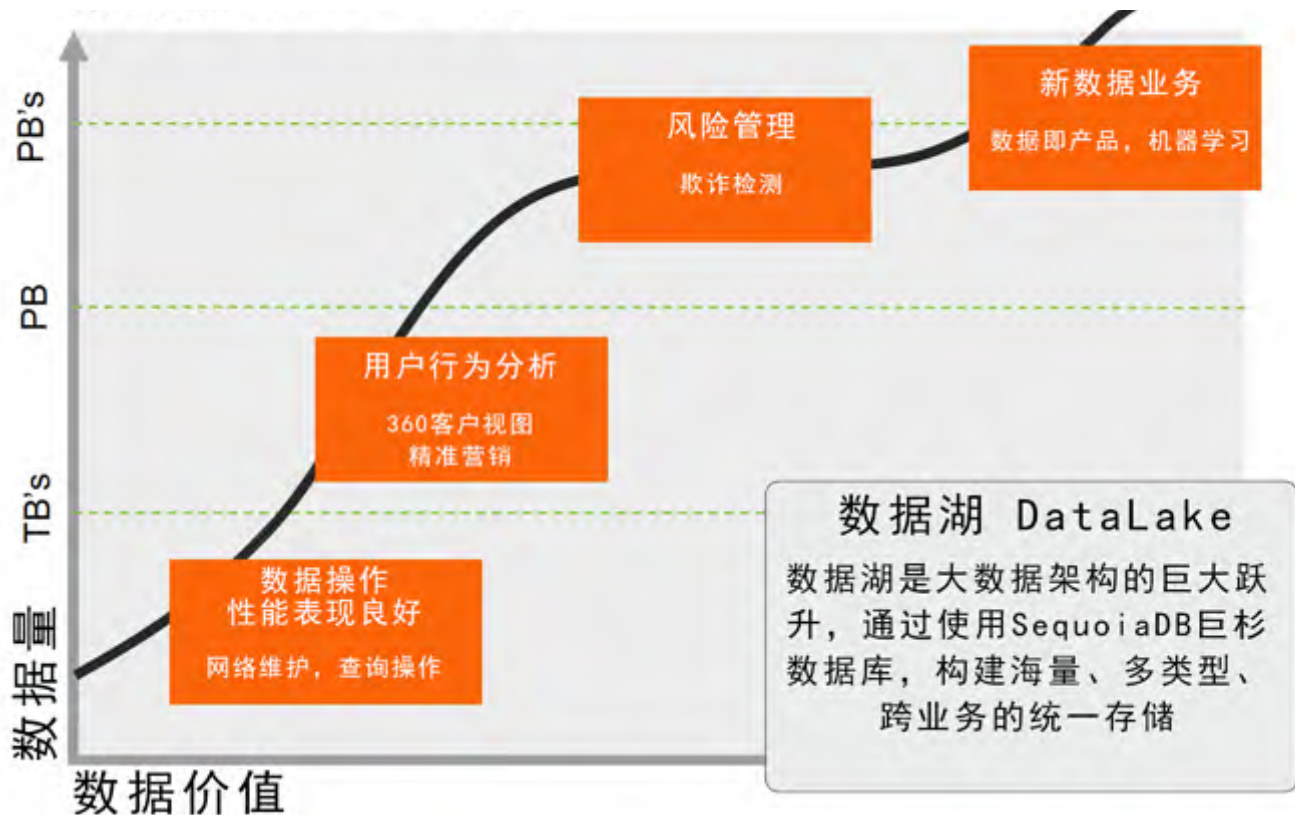
SequoiaDB

大数据数据湖 (Data Lake)

传统的数据集市，为了优化数据分析的性能，只会存储一部分的属性并且只操作已经聚合的数据。

而“数据湖”则是将**全量数据的所有属性**都进行存储，尤其是当你并不确定从什么角度去分析和使用这些数据时

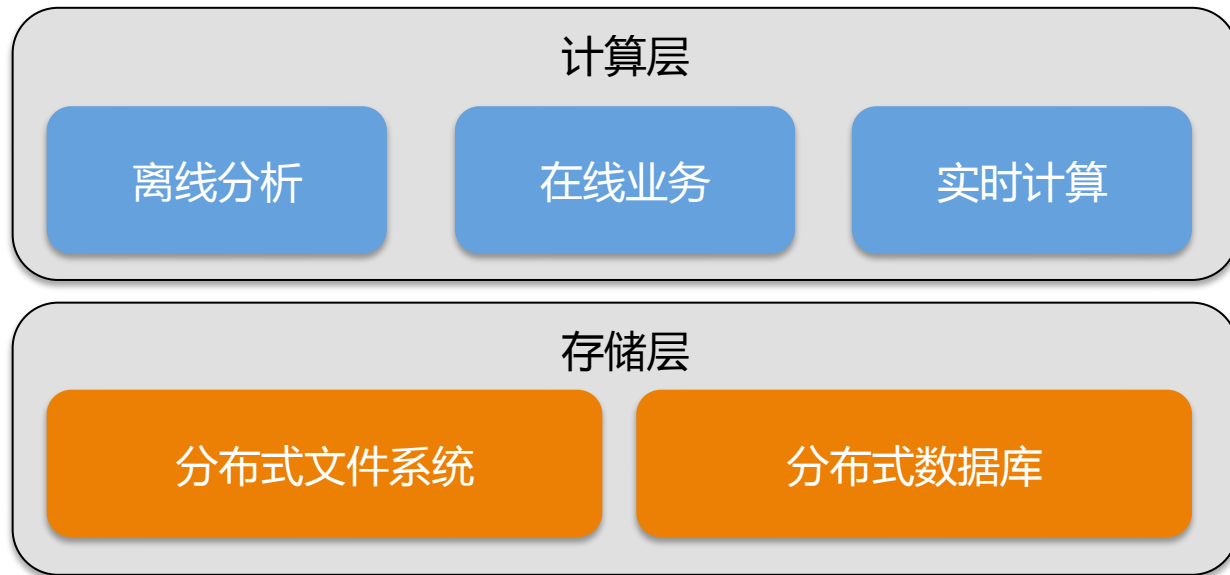
数据湖实现大数据价值 (Data Lake)



数据湖的优势



大数据技术框架划分



大数据时代的数据库

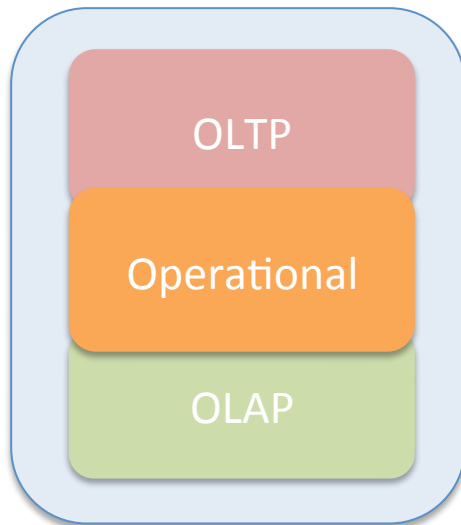
大数据时代

---OLTP，OLAP外的新需求

大数据应用场景，催生出了传统OLTP和OLAP外的新需求就是 **实时数据操作（Operational）**。

主要的技术需求包括：

- 海量数据的存储管理
- 高并发实时查询
- 高性能的读写



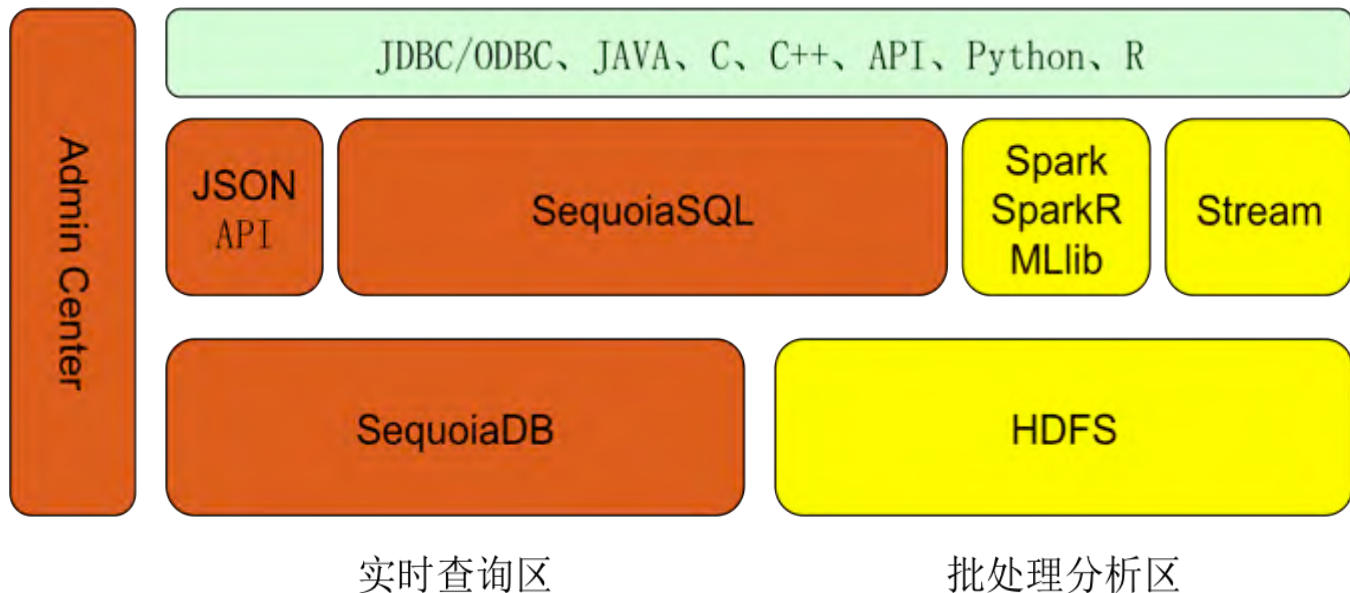
大数据基础软件的发展：殊途同归



分布式数据库的未来

- 传统关系型数据库
 - 成为存量市场，逐渐萎缩
- NoSQL
 - 开始支持ACID与关系模型，在提供自身特有特性的同事提供标准SQL接口
- SQL-on-Hadoop
 - Hadoop成为分布式文件系统与分布式调度框架，MR被完全淘汰，原生SQL引擎一统天下，Spark负责流处理与分析挖掘
- NewSQL（新型分布式数据库）
 - 概念逐渐被泛指新型分布式数据库，传统NewSQL、新型的支持SQL接口的NoSQL与SQL-on-Hadoop都可以被概括在NewSQL的范畴之内

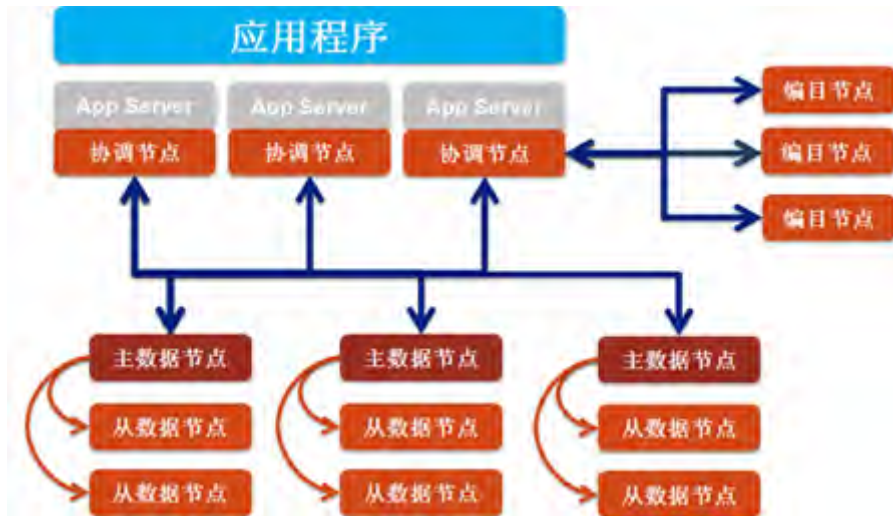
NewSQL数据库的产品定位



NewSQL技术特点

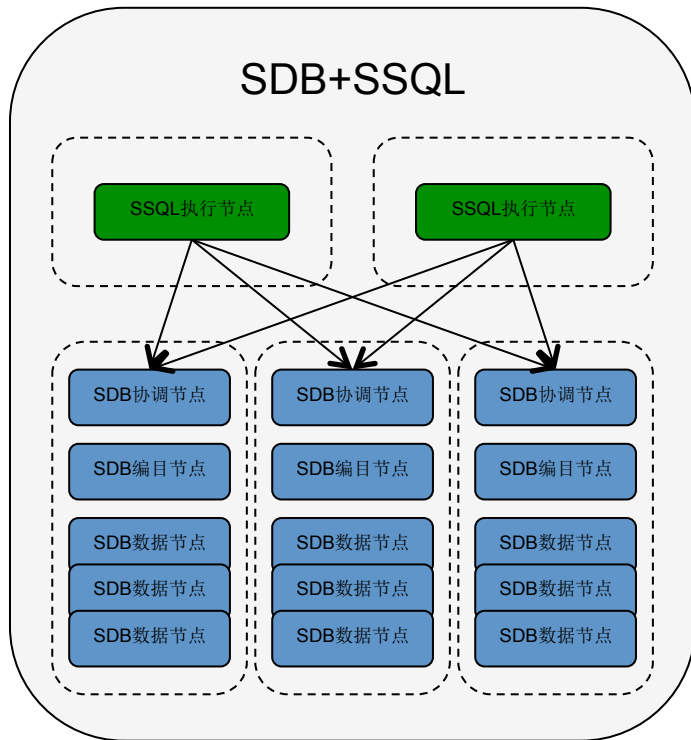
分布式架构与扩展性

- 分布式 Share-Nothing架构
- 主-备 高可用机制
- 线性水平扩展



SQL支持

- SQL支持：ANSI SQL 2003 标准，OLAP扩展，标准 JDBC/ODBC支持
- 支持多种分区方法及多级分区
- 多种UDF（用户自定义函数）语言支持：java, python, c/c++, perl, R等
- 动态扩容：动态按需扩容，按照存储大小或者计算需求，秒级添加节点。



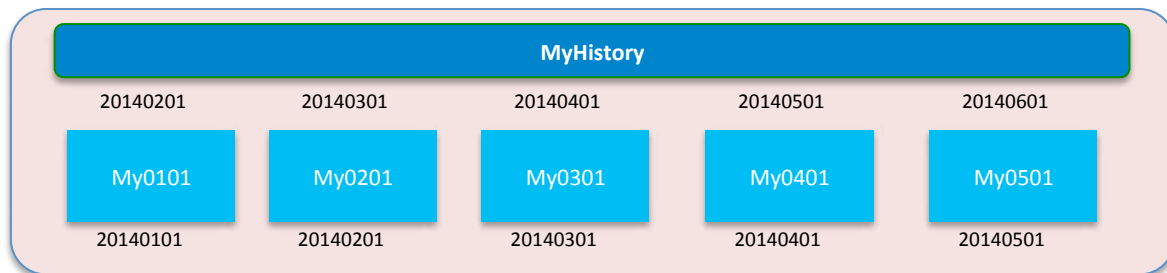
数据分区/主子集合机制

- 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
- MongoDB和Hbase都没有提供类似的主子集合机制

```
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" }}
)
```



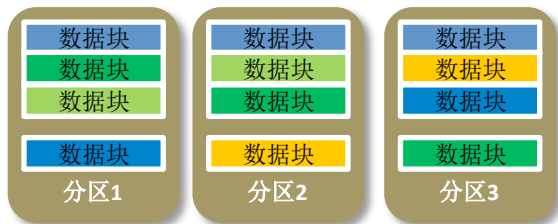
双存储引擎

记录/块存储

- SequoiaDB分布式块存储

文件按照LOB处理

- 自动按照64/128KB的数据块进行切分，放在不同分区存储
- 使用DIO避免二进制数据占用文件系统缓存
- 并行处理
- 与GridFS相比不占用内存
- 与HDFS相比不存在NameNode限制



- 其他实现方式

1. 关系数据库+文件系统地址

问题：文件条目受关系型数据库的限制，比如超过3亿条后性能下降

2. MongoDB GridFS

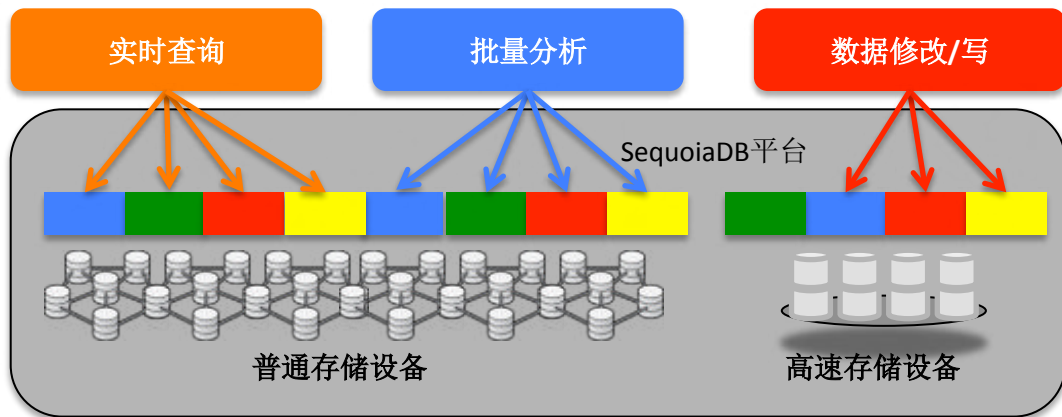
问题：MMAP机制把文件放进内存，与索引等关键数据争用内存

3. HDFS

问题：受 NameNode限制无法处理大量的小文件，分配64MB存储浪费

读写分离

- 数据在多个分布节点内可以自动分片和复制，可以定制数据分布策略.保证多类型业务可以运行在同一平台上，比如：
 - 冷/热数据区分离
 - 读/写分离
 - 写交易的“强一致性”和“弱一致性”分离
 - 查询/批量分离



分布式事务

一致性因素	读主节点	读备节点
$W=N$	强一致，无数据丢失，事务操作符合传统关系型数据库模型	强一致，无数据丢失，读操作受节点分布影响会有脏读时间
$W! = N$	强一致型，极端情况下可能发生数据丢失，事务操作符合传统关系型数据库模型	最终一致，可能发生读取不到写入的数据，极端情况下可能发生数据丢失

持久性因素	备节点响应策略	数据可靠性
物理同步	数据在备节点写入事务日志	全部节点宕机不会造成数据丢失，但会损失性能
逻辑同步	数据在备节点处理但未写入事务	主备节点同时宕机可能会造成数据丢失，数据查询强一致
半同步	数据在备节点成功接收但还未处理	备节点宕机可能会造成数据丢失，最终一致性
异步	数据不需要被备节点感知	主节点宕机可能会造成数据丢失

MongoDB和Hbase 不支持SQL访问，其API也没有提供事务机制。

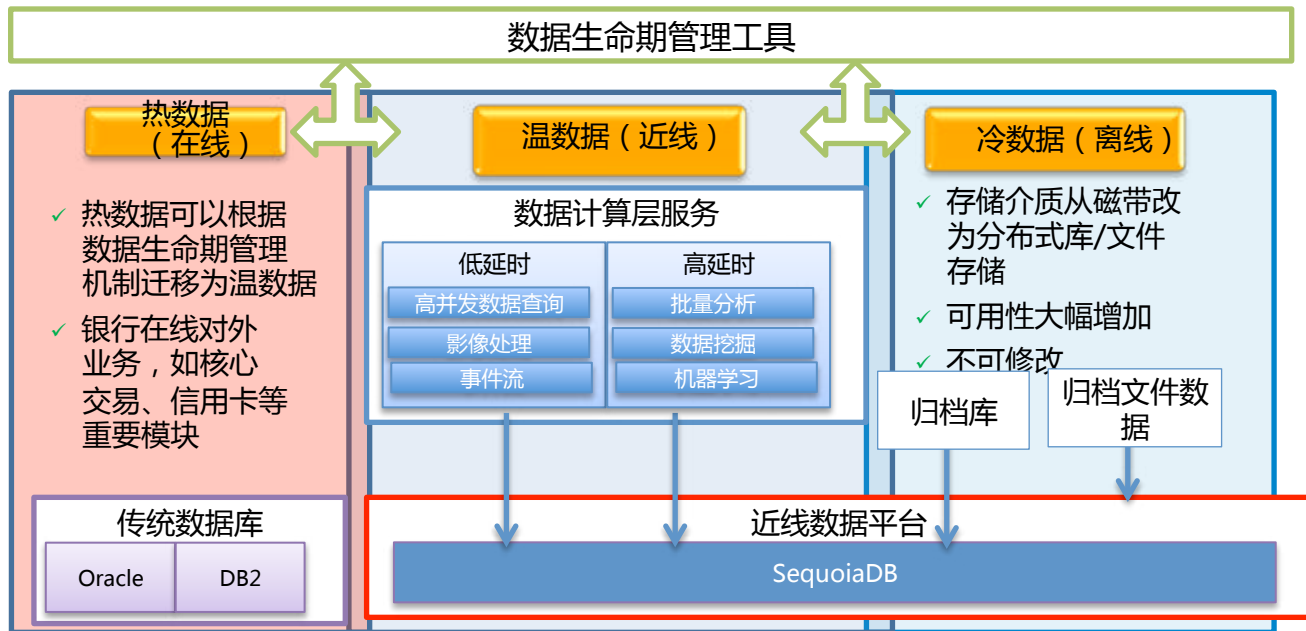
NewSQL应用案例

1. 金融行业近线数据平台

近线数据服务平台体现数据湖业务价值

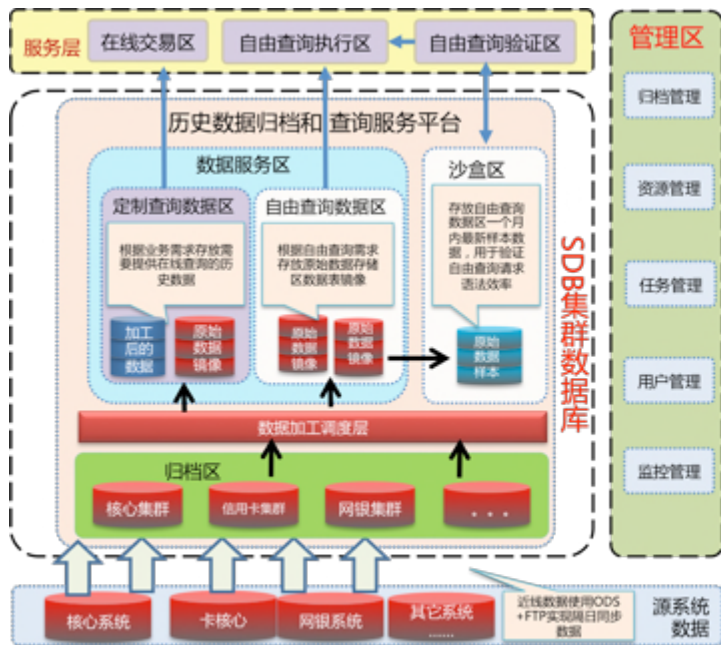
- 离线数据近线化
 - 业务系统数据的统一归档
 - 提供冷数据对最终用户的服务（例如ECIF、回单查询）
- 自由查询、自助报表
 - 面向监察室及运维人员开放统一的数据查询入口
 - 面向公检法的司法查询系统
 - 面向行内用户与分行IT人员，提供自定义数据查询与报表服务功能
- 生产系统瘦身
 - 数据仓库、ODS瘦身
 - T+0用户实时资产视图

近线数据平台定位



NewSQL数据库构建近线数据平台核心价值

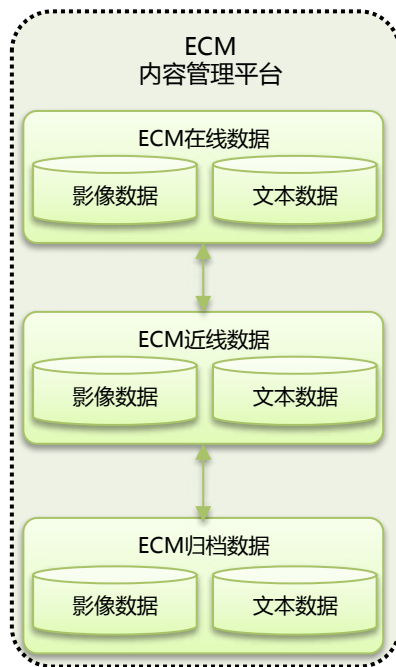
- 海量的离线数据的近线化
- 近线数据平台存储原本离线与近线数据，不涉及已有在线业务
- 对已有业务不进行替换，而是补充
- 达到离线数据近线化、与近线生产库瘦身两大目的
- 分布式架构提升性能，节约成本



2. 企业内容管理/影像平台

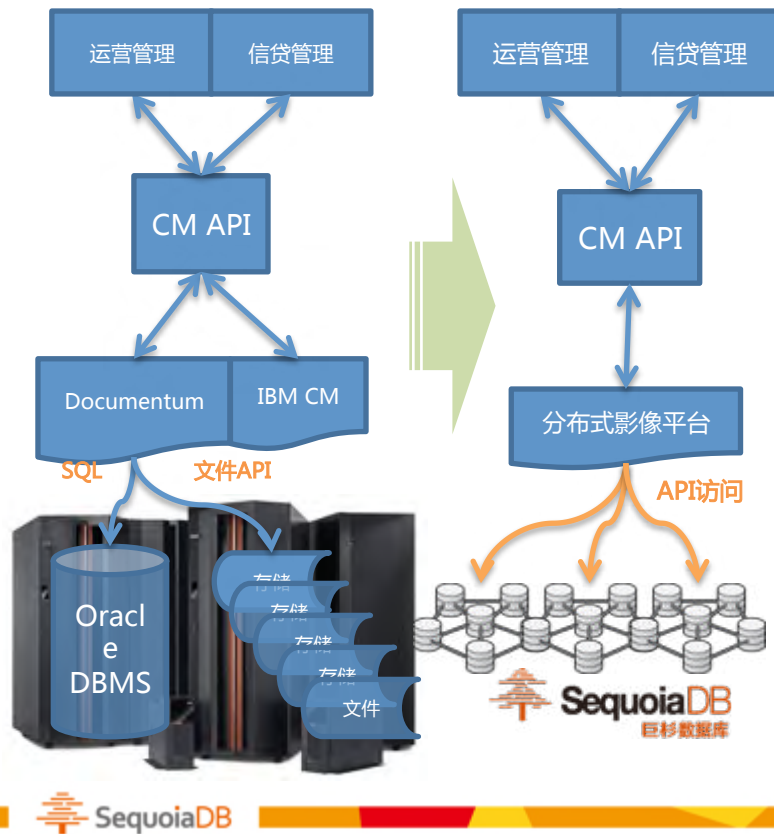
ECM影像内容管理

- 存储票据、影像、视频等非结构化数据区域
- 提供数据分层存储机制
 - ECM在线数据：提供热数据存储服务，使用高端PC服务器保证读写速度高效稳定，支持大规模并行访问和读写能力
 - ECM近线数据：提供中期数据存储服务，以只读业务为主，对修改操作支持但性能欠佳
 - ECM归档数据：提供数据的归档服务，在SDB中归档数据也可以进行访问，只是使用相对低配置服务器导致不以性能为优先考虑点
- 提供在线、中高并发的内容数据处理
- 提供15年影像视频数据保留期限

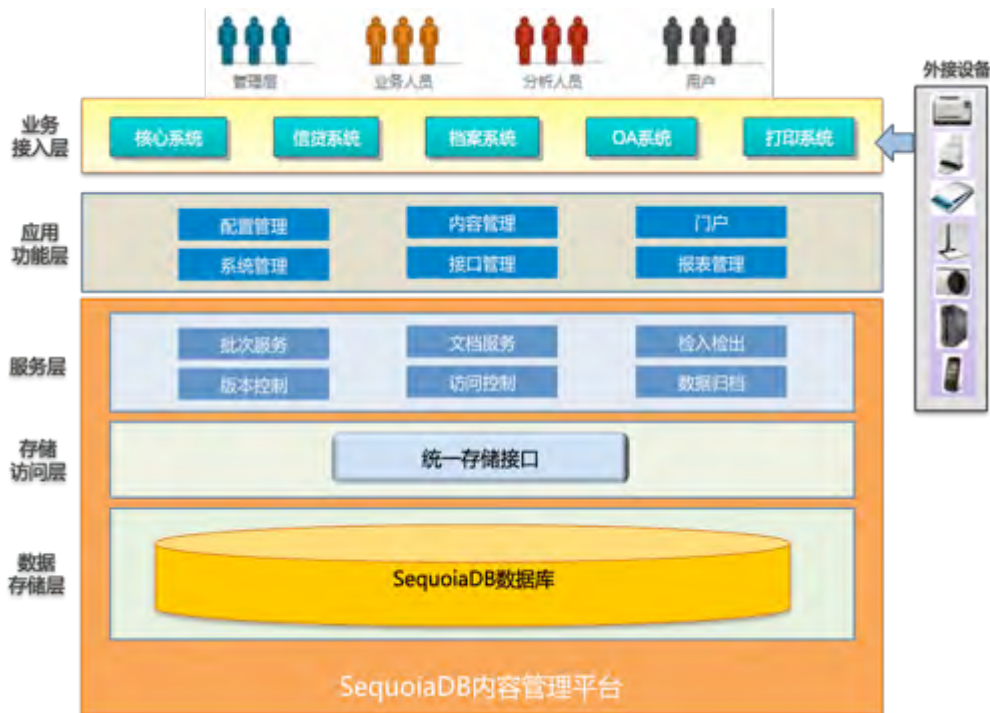


NewSQL数据库构建 ECM内容管理平台

- 双引擎机制，文件和元数据同时存储
- 分布式架构提供容量的灵活扩展，提高性能节约成本
- 海量非结构化数据的实时查询、调用，前台业务的无缝对接



SequoiaECM逻辑架构



- 多渠道内容管理
- 文档管理
- 用户鉴权身份验证
- 检入检出机制
- 文档多版本控制
- RESTful API接口
- 对象存储接口
- 多平台支持

3. 360用户视图/数据视图

政府大数据平台

某市电子政务中心利用SequoiaDB，作为统一的数据湖，将每个市民的身份证作为唯一标识，将全市1000万+市民以及超过40余个委办局的数据打通，做到针对市民的全维度视图。

在该系统中，SequoiaDB利用其分布式架构与对象存储能力，将每个市民的全维度数据形成一条统一的记录，不同委办局的维度在该记录中以嵌套字段的形式体现，大大方便了应用程序的开发和管理。



互联网旅游平台数据撮合

途牛旅游网，使用SequoiaDB巨杉数据库支撑动态资源和动态报价等业务，通过应用于“价格中心”、“资源系统”、“酒店系统”等多个业务核心部门，解决了海量多类型数据的存储以及处理这个互联网企业普遍面临的技术瓶颈，使得多个核心系统计算量从去年同期的每天1亿次，增加到今年每天100亿次以上。



4.物联网智能制造/智慧交通

NewSQL for IOT

- 分布式架构，应对物联网快速增长的设备数据
- 多类型存储，日志、图像等非结构化、半结构化数据的存储
- 高性能高稳定，为物联网/智慧交通提供了实时的数据存储管理和调用支持

物联网与智能制造

如今的IT已经逐渐由“人产生数据”向“设备产生数据”过度。对于设备的制造企业来说，传统的IT主要围绕ERP、库存、上下游渠道、CRM等系统构造，而对于设备本身产生的数据则没有过多关注。但是，对于新一代物联网设备，其产生的数据具有量大、繁杂、多样等特性，使用传统数据存储与处理技术很难有效地从中挖掘出价值。



在巨杉某运营商用户中，SequoiaDB作为智能网关的后台数据管理平台（DMP），承载智能网关设备中软件日志的采集、分析与挖掘功能，为家庭用户更安全、稳定、高效的网络环境提供了基础保障。

智能交通监控

在政府与交通行业，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。



5.互联网金融业务底层数据平台

互联网体系下的金融模式

- 独立于核心银行之外的新型服务模式
- 在商业与技术上探索新的渠道对金融行业所带来的转变
 - 渠道由线下转移到线上，由人与人的接触转变为机器与人的接触
 - 信息的收集和发送从人转向软件
 - 用户体验要求的提升
- 新的业务模式
 - 远程开户、虚拟柜员、网上信贷、非金融服务、个人与小微征信
- 新的技术模式
 - 原本由人做出的决策，现在需要由机器完成
 - 用户行为数据与第三方数据是新模式的核心

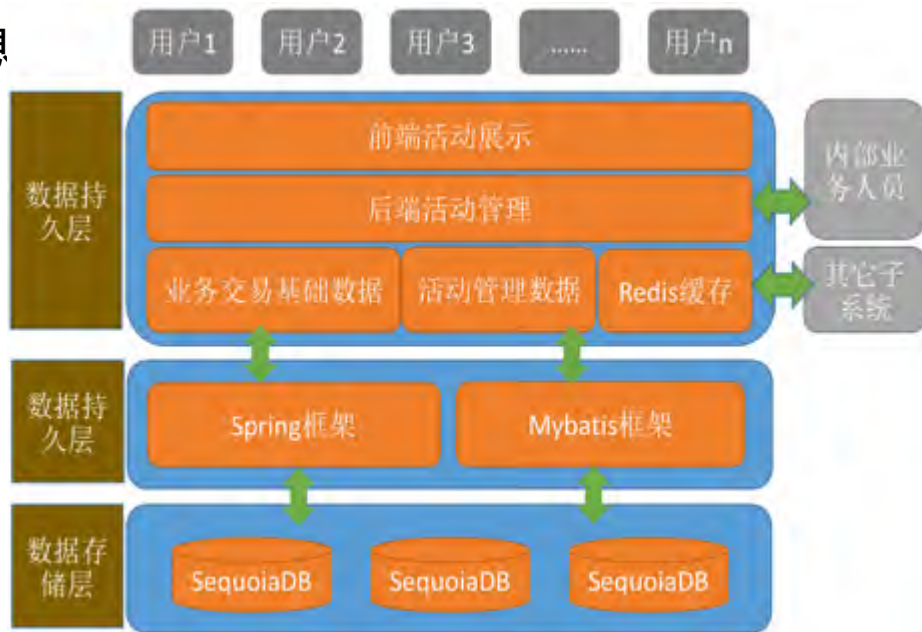
NewSQL数据库 助力 互联网金融业务

- 分布式架构，灵活容量扩展，应对飞速增长的互金业务数据
- 灵活存储结构，结构化非结构化数据同时存储，适应互金数据系统的多种数据处理需求
- 兼容企业级应用，SQL等企业级功能，完美适应已有应用
- 高性能的存储管理，高速响应互金业务的高速发展
- 分布式高可用，保证互金数据的安全稳定



案例：NewSQL应用于电子银行

- 优惠券、卡券信息
- 基础交易数据
- 活动管理
- 渠道管理
- 精准营销
- 日志采集





关注巨杉公众号
试用最新2.6社区版

www.sequoiadb.com
Sales_support@sequoiadb.com
400-8038-339