SDCC
中国软件开发者大会
SOFTWARE DEVELOPER CONFERENCE CHINA

CSDN
全球最大中文IT社区

海量并行SQL(RapidsDB)在大数据深度学习中的实践

-刘睿民
ISO SC32 SQL Standard International Expert
柏睿数据董事长兼CTO

- 支持 ANSI 2011 SQL standard 以及 SQL OLAP 函数。包含了联邦（Federation）的支持，可以对很多数据源如 DB2、Oracle、Teradata、Hbase 等进行联邦访问。联邦功能允许用户在同一个 SQL 语句内给各个关系型数据源发送分布式的请求。

- 工作节点允许将大的数据集 spill 到磁盘，允许 Big SQL 处理超过最大可用内存大小的数据结果集。同时允许实时处理及分析来自多个外部来源的数据。

- 具有基于统计信息驱动的优化机制，Big SQL 通过维护表级别、分区级别和列级别的丰富统计信息集合，可以帮助查询优化器优化查询计划，支持 Nested loop join/Sort-merge join/Hash join 等。

- 通过查询重写机制（query rewrites）自动决定一个查询的优化方法以便减少查询所需的资源，比如如果相同的表达式在同一个查询中出现多次，查询将只计算该表达式一次并会在多个地方重用该结果值。甚至如果查询多次引用相同的表，则只对该表进行一次扫描满足所有引用需求。

RapidsDB MPP SQL Engine

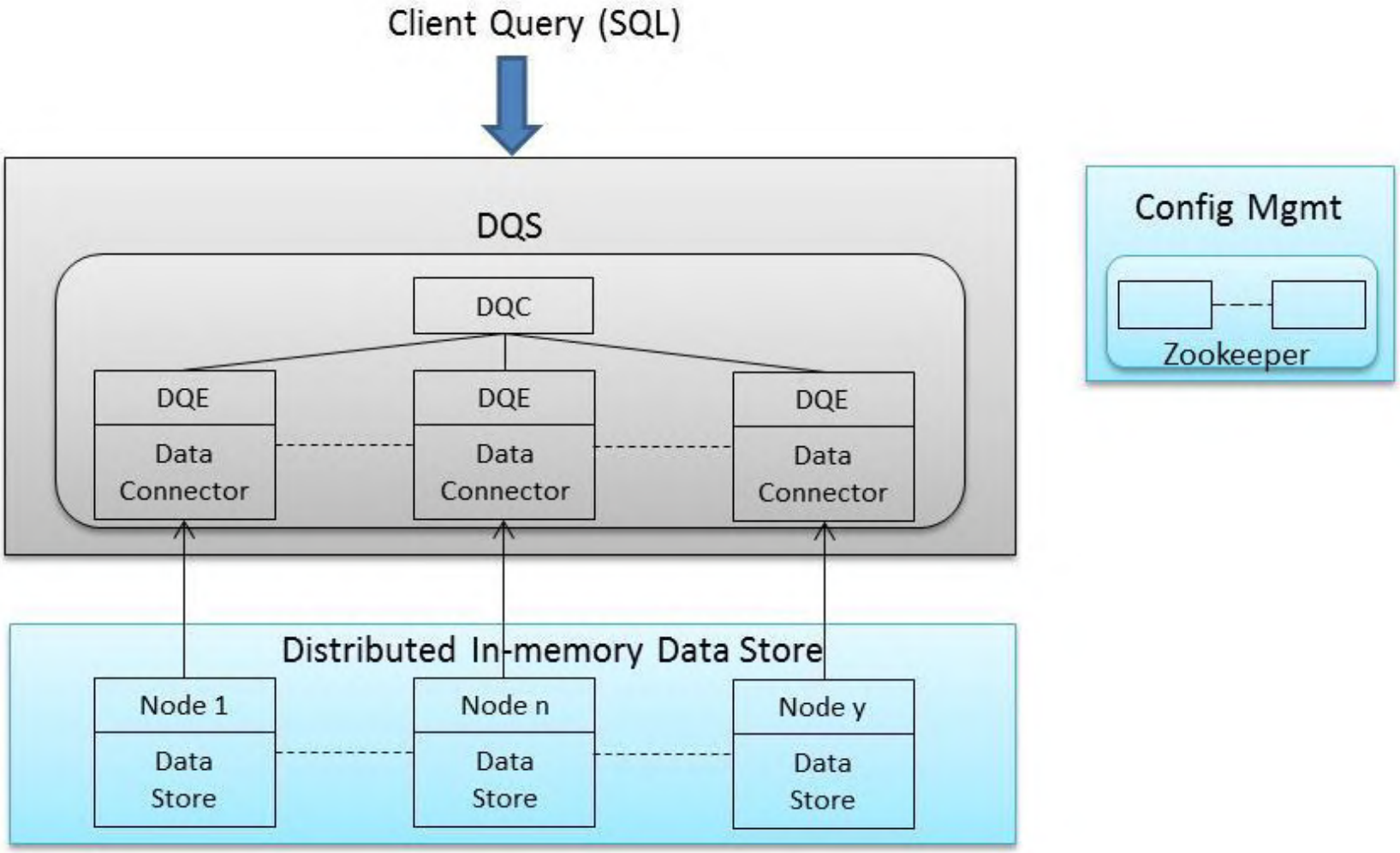
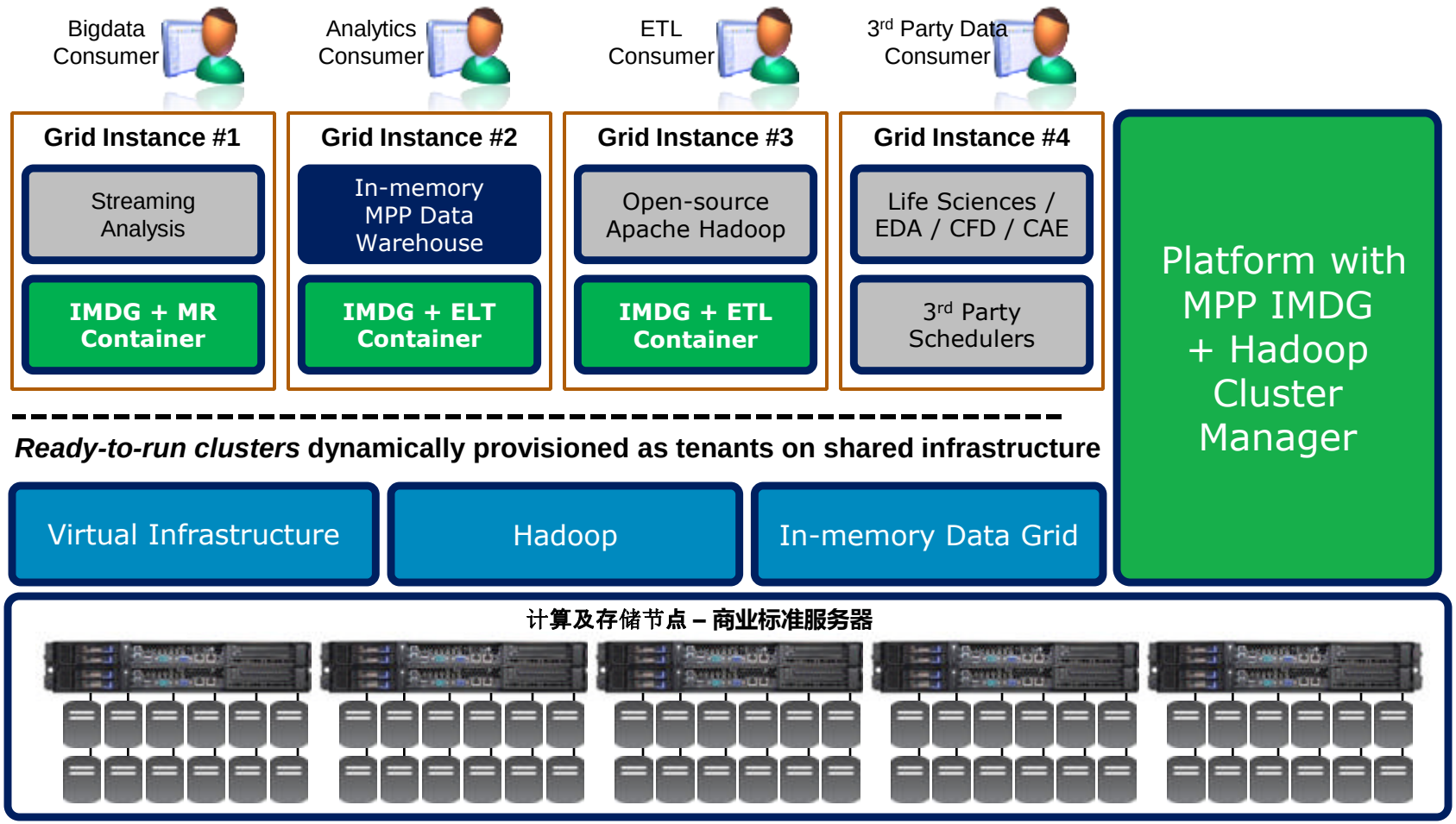
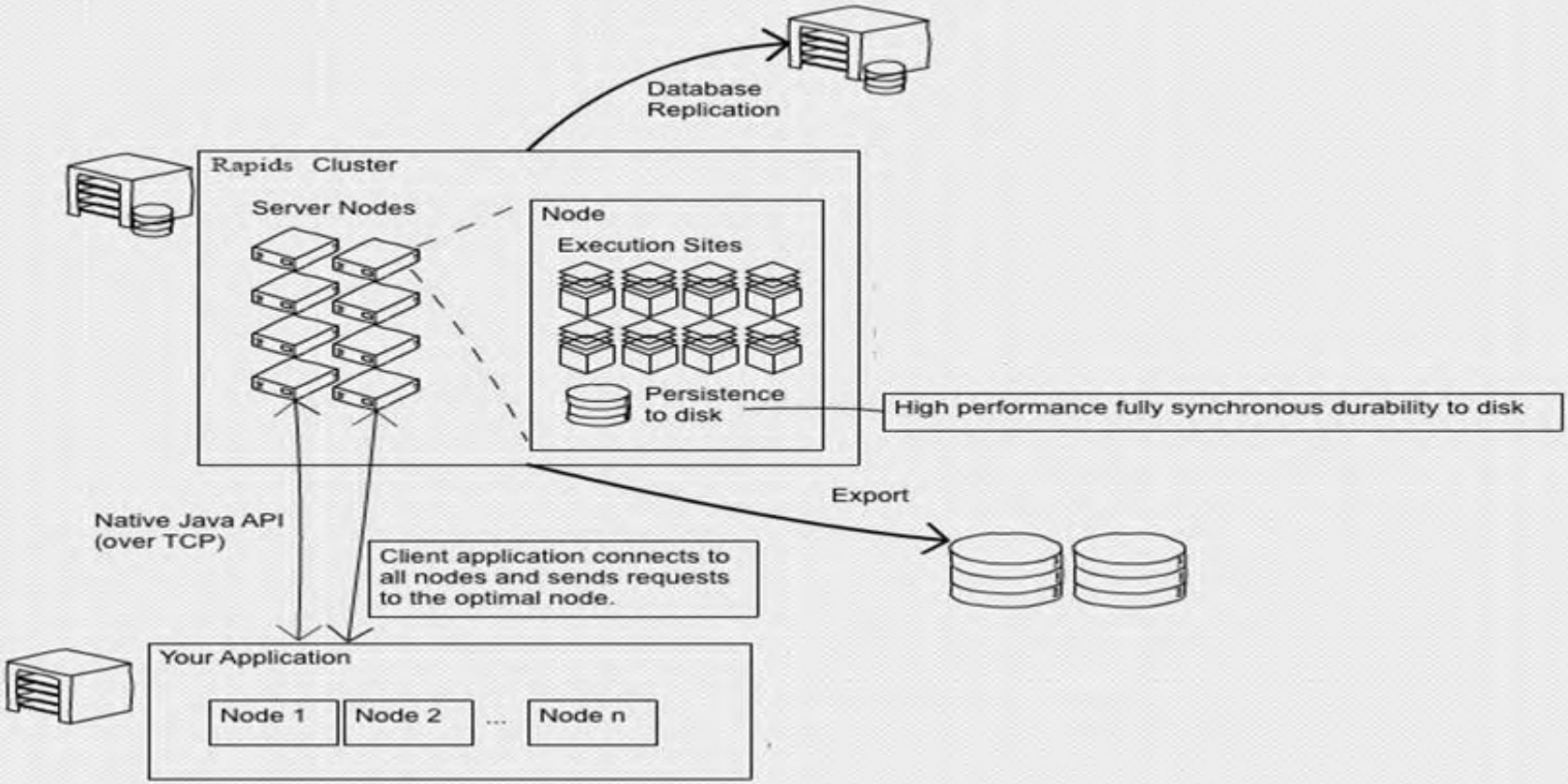


Figure 1. RapidsDB architecture

核心架构 (Logical View)

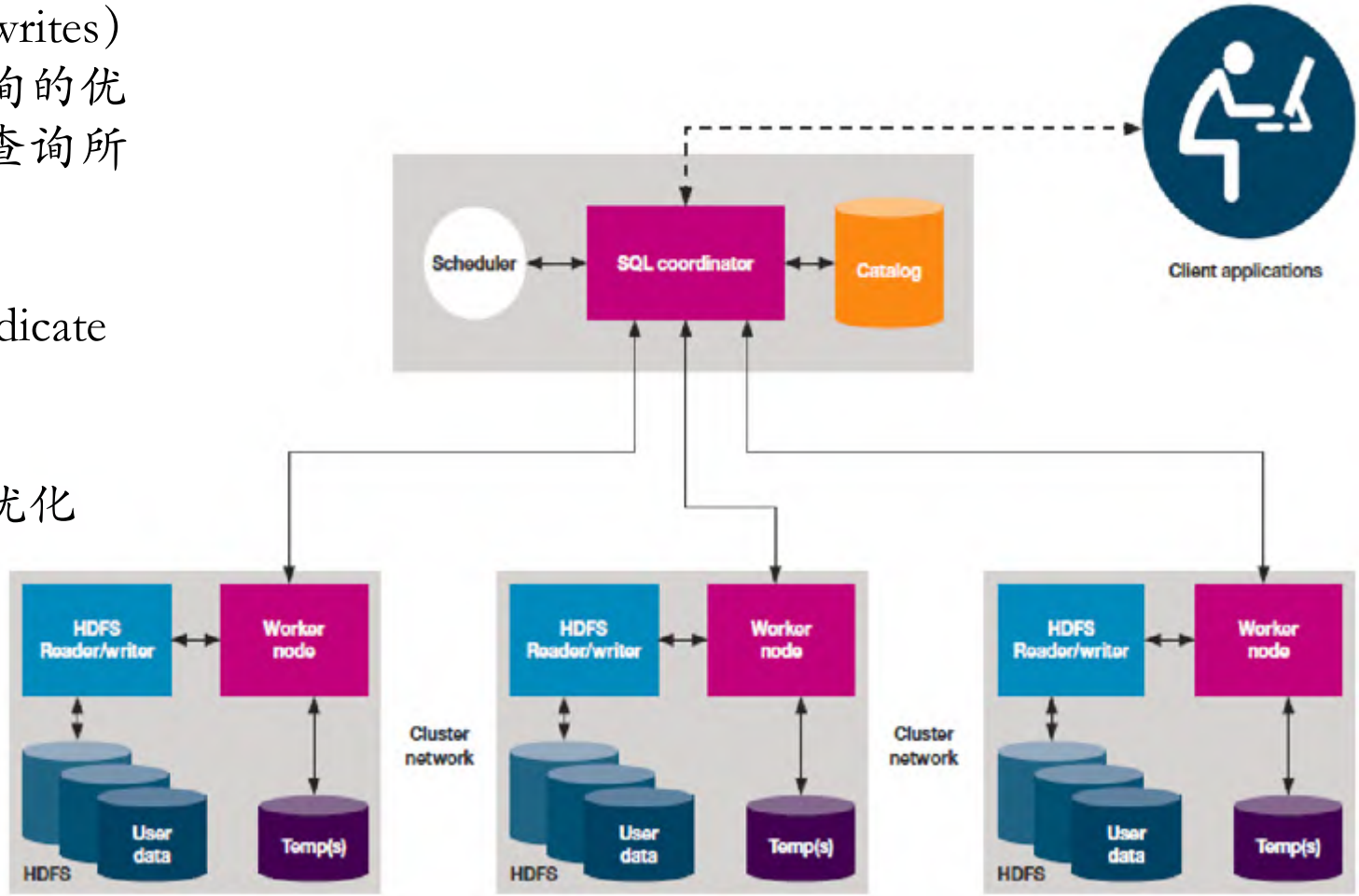


RapidsDB MPP SQL 物理架构

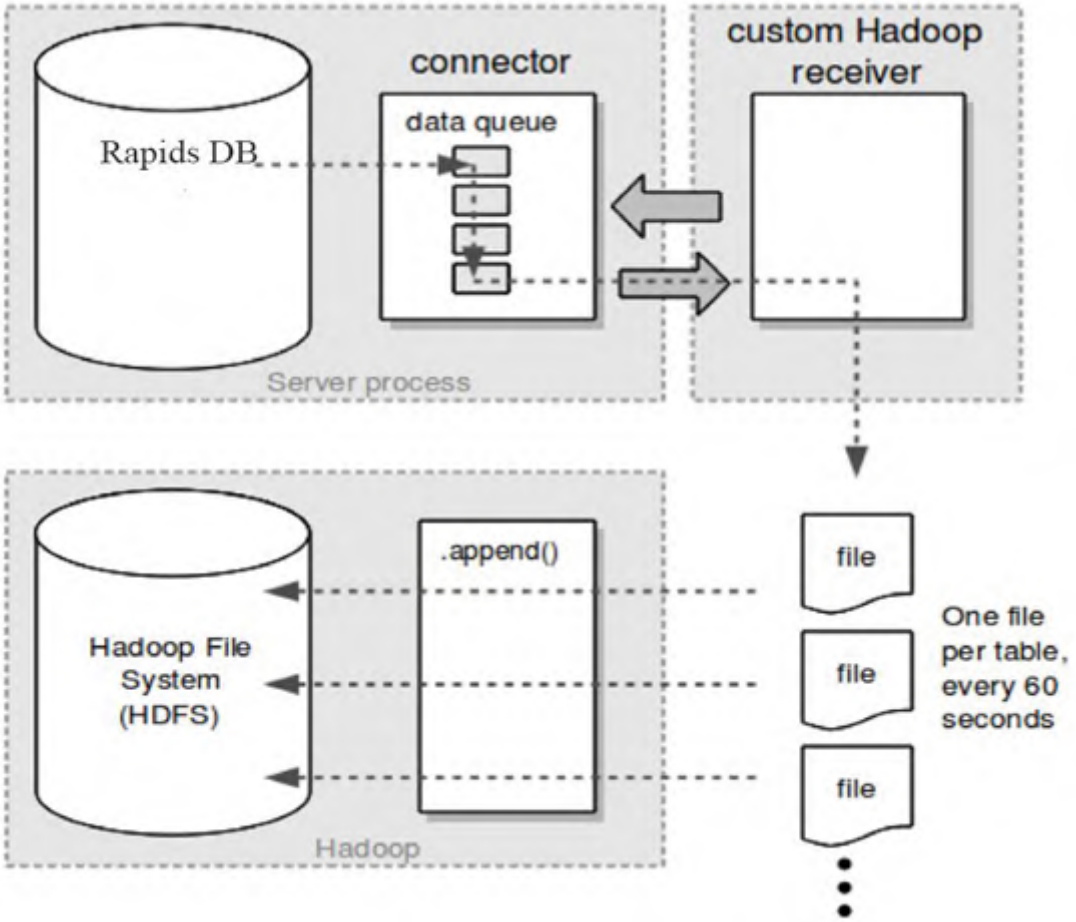


MPP SQL引擎优化

- Big SQL 通过查询重写机制（query rewrites）自动决定一个查询的优化方法以便减少查询所需的资源
- 谓词下推（Predicate pushdown）
- 统计信息驱动优化（Statistics-driven optimization）



Big SQL On Hadoop的无缝集成



特点：

1、数据流实时导出。

2、 数据流输出格式多样性。

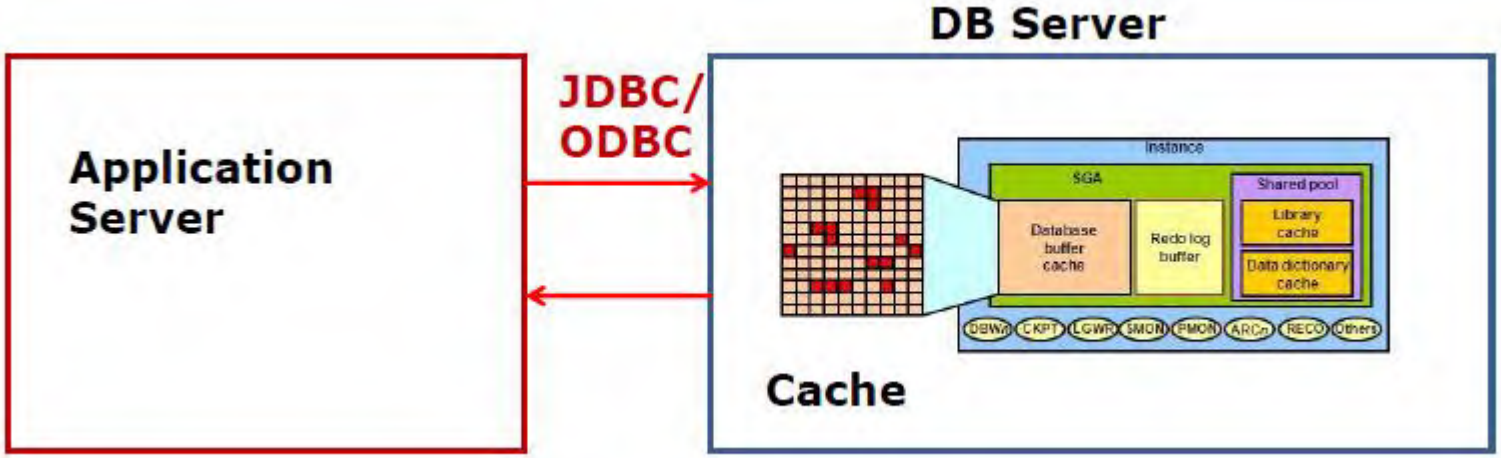
3、数据流内存溢出保护。

4、自定义数据流导出接收器，方便用户扩展

5、批量写HDFS，写入速度快

6、今后会支持使用Stream直接操作HDFS中的数据

内存对象并行访问及处理

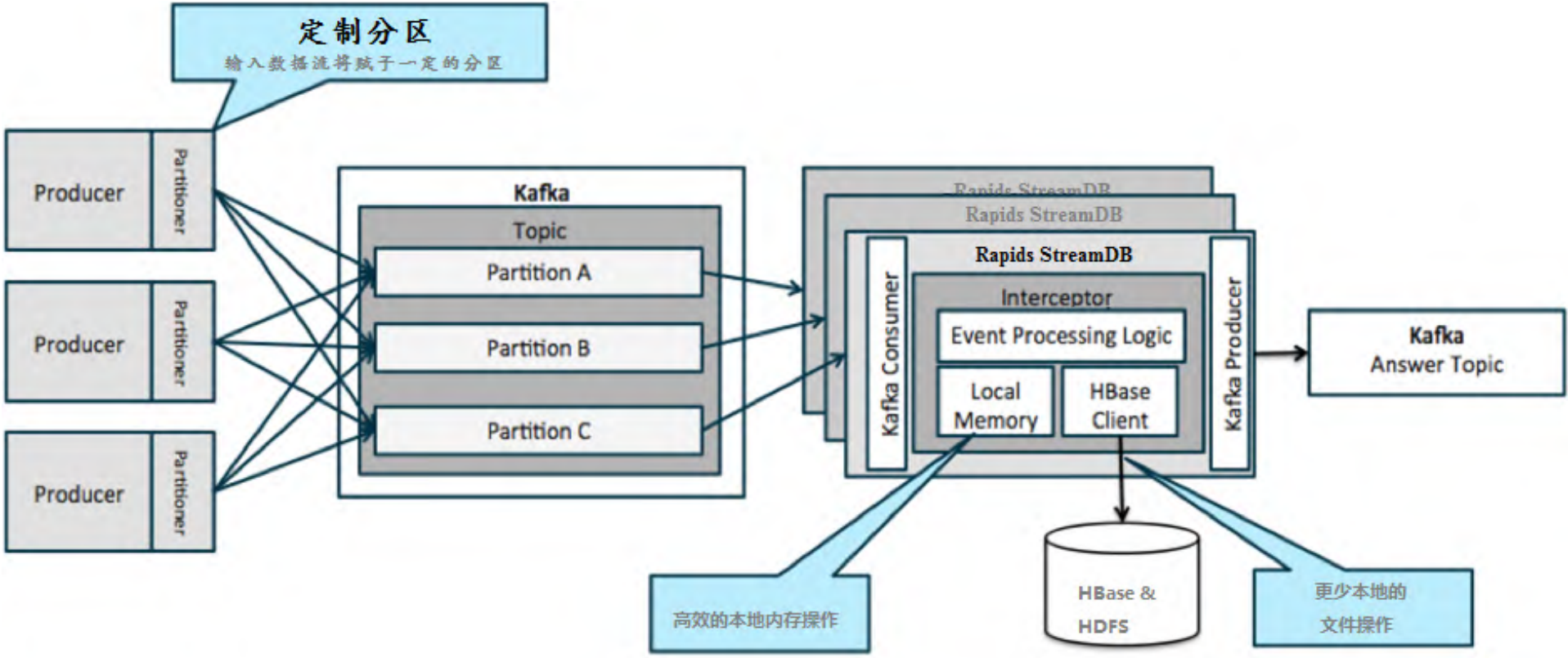


远程TCP访问带来非常昂贵的系统开销. 而内存访问的延迟在几分之毫秒. JDBC/ODBC 调用带来的延迟会高达100-1000倍于毫秒。

在App与DB间的通讯负载可以轻易的达到整个服务器端的30%。

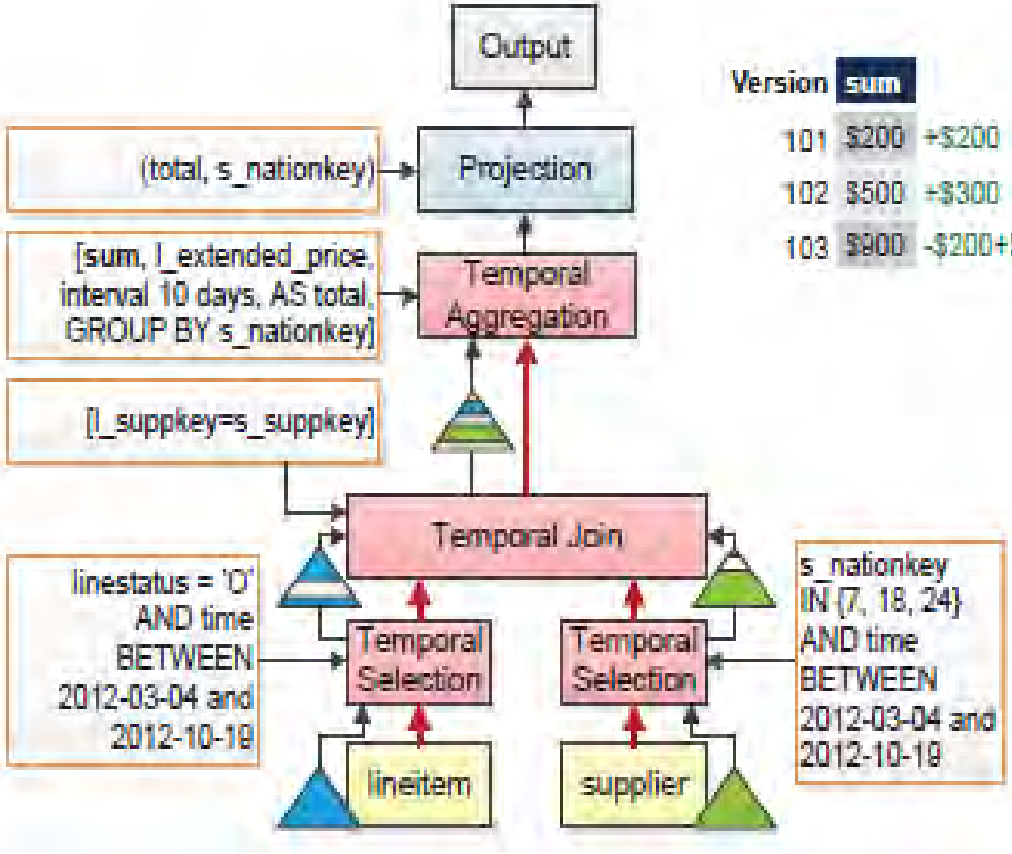
日志分析实践

结合时间序列功能的架构

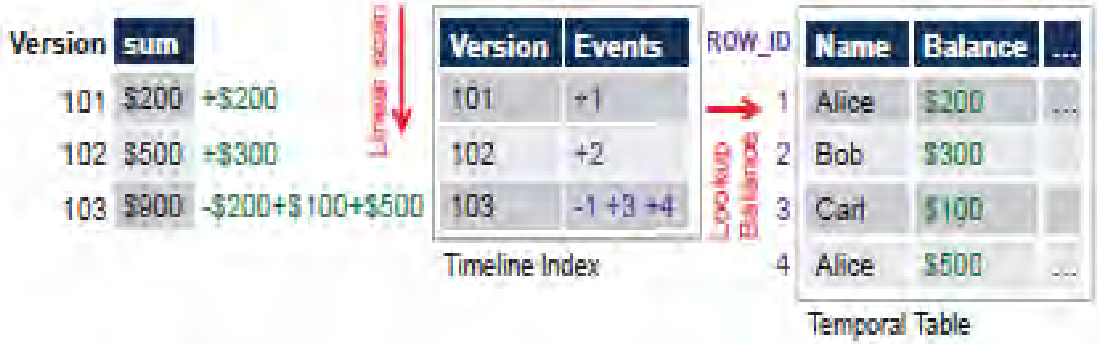


时间序列功能

时间序列数据的JOIN

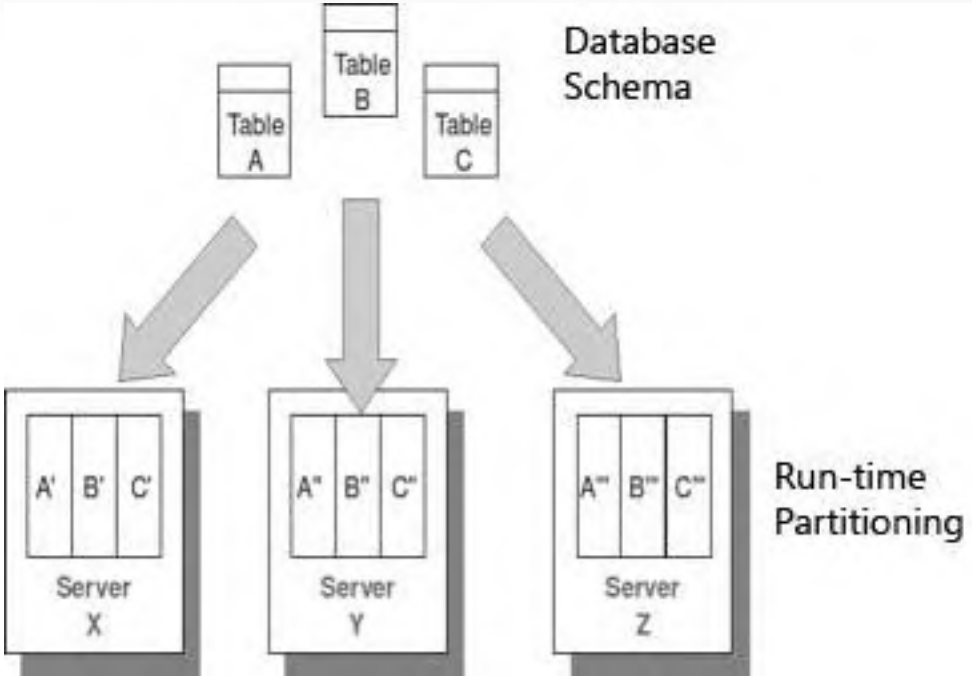


时间序列数据的SUM



数据分区执行机制 (Parallel Partition)

- 分区表 (Partition Table)
  - 数据量较大或写较多的表
  - 每个分区绑定到每台服务器的一个CPU核心
  - 客户端无需担心数据的位置
- 复制表 (Replica Table)
  - 通常为较小或读操作较多的表
  - 复制到每台服务器较少跨节点操作
- 作为开发者，只需指定每个分区表的分区键，RapidsDB会自动根据键值将数据分区。



- 分区表被自动切分到不同分区
- 每个分区绑定一个CPU核心(Core)
- 复制表在每台服务器有一个副本

共同发展，
为行业大数据
的存储和实时
分析贡献力量

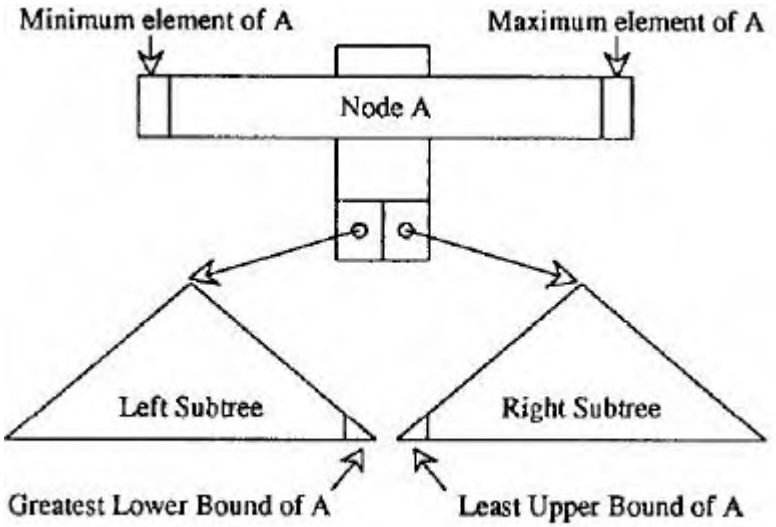
柏睿数据科技（北京）有限公司
期待成为您的长期合作伙伴！
感谢您对柏睿数据的支持！

内存索引的并行化(Index Parallelism In Memory)

- 1986年Lehman & Carey的论文关于T-Tree性能指标就已经显示，T-Tree在内存内的memory insert, search, range scan, and delete性能至少比B Tree快20%－50%。

- T Tree 的重平衡表现类似 AVL Trees, 会产生滚动想象，实现有比较高的难度。

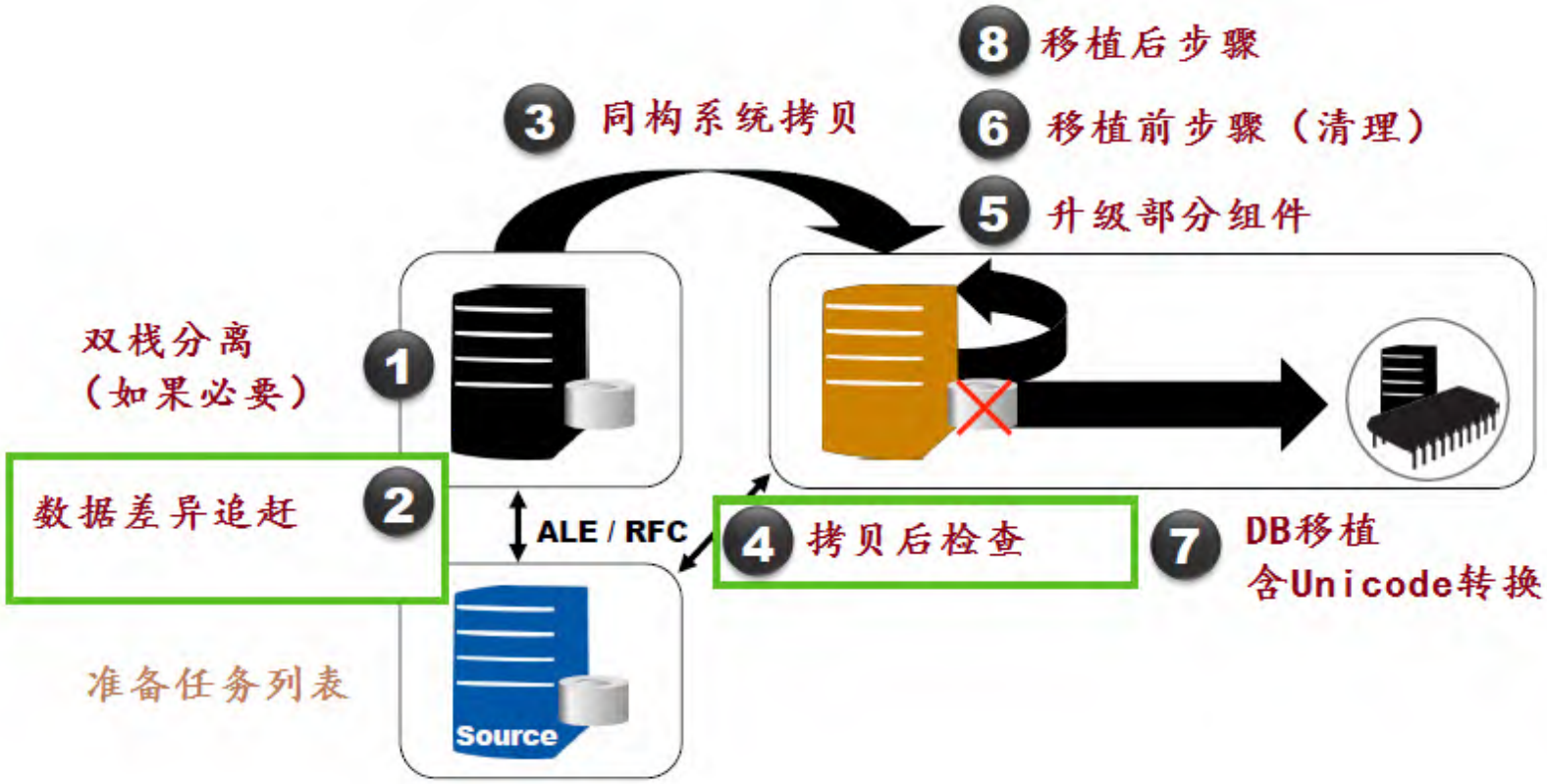
Lehman, Carey: VLDB'86, A Study of Index Structures for Main Memory Database Management Systems



ISO/IEC 9075:SQL 2011的支持

- ISO/IEC 9075:SQL 2011的支持，未来SQL 2013
- SELECT, INSERT, UPDATE, DELETE + Stream
- SQL支持AVG, COUNT, MAX, MIN, SUM等SQL函数列表
- SQL LIKE
- 子查询
- 视图
- JSON值
- SQL支持仍在不断增加
  - 不支持的功能可通过Java存储过程实现
- 针对在线分析性事务处理优化
  - 分析型事务引擎

日志分析步骤


8 移植后步骤
3 同构系统拷贝
6 移植前步骤（清理）
5 升级部分组件
双栈分离
（如果必要）
1
数据差异追赶
2
ALE / RFC
4 拷贝后检查
7 DB移植
含Unicode转换
准备任务列表
Source