

DTCC

数 / 造 / 未 / 来

第十二届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2021

2021 年 10 月 18 日 - 20 日 | 北京国际会议中心

浪潮云溪数据库3.0

陈磊
浪潮云溪数据库研发副总经理兼产品负责人



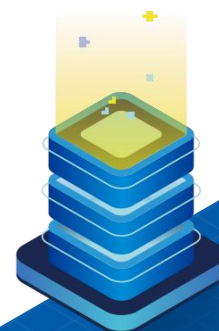
目录



一、浪潮云溪数据库简介

二、云溪3.0核心特性

三、云溪应用案例



一、浪潮云溪数据库简介

新一代互联网、物联网和工业互联网的数据管理基础设施



产品内涵

数字经济的发动机

- 纯分布式架构，异地多中心单一分布式集群
- 海量数据存储、加载与分析
- 高可用、弹性扩展



价值主张

基座可靠

- 技术先进、产品可靠、安全易用
- 与用户共同成长



云原生

- 扩展性强，弹性伸缩
- 高可用，故障转移
- 易运维，快交付



分布式HTAP

- 同时处理OLTP和OLAP
- 分布式事务，数据强一致
- 无须分库分表



治理

- Oracle兼容替换
- MySQL兼容替换
- 微服务支持、分级权限管理

红旗®
Linux

PHYTIUM 飞腾

Kunpeng

银河麒麟
KYLIN

中标麒麟
NeoKylin

统信软件
UNIONTECH

一、浪潮云溪数据库简介

三年，三个阶梯

- 进一步提升Oracle/MySQL/PG语法及协议兼容度
- 等保三级安全支持
- 准内存引擎/列存引擎/算子等优化
- 多中心分布式部署增强
- 可视化运维工具/外围生态工具支持

 2020.3

- 2019年，浪潮研发立项，确定基于Spanner理论，采用纯分布式技术实现分布式数据库
- 2020.3发布1.0版本，具备分布式数据库的基础功能，具有强一致、高可用、分布式、多中心等特性

 2022.2

3.0
云溪

2.0
云溪

- 兼容传统数据库的语法、函数
- 支持游标/分区/索引等传统数据库业务
- 完成国产软硬件的兼容
- 公有云版本上线，充分发挥云原生特性

 2021.3

1.0
云溪

一、浪潮云溪数据库简介

云溪3.0 物联时代的云原生数据库



- “工业”属性的互联网数据形态多样
- 存储于关系数据库/时序数据库的结构化数据
- 半结构化或非结构化数据



- 需要数据分类分级安全管理的解决方案
- 提供标准接口的动态生成及现有接口的注册
- 数据平台需要支持多计算和存储引擎



- 成本敏感，云原生带来硬件接耦
- 弹性伸缩，灵活应对大促秒杀抢票等应用场景



- 数据资源缺乏整合，难以有效管理和利用
- 数据质量参差不齐，标准化和规范化有待统一



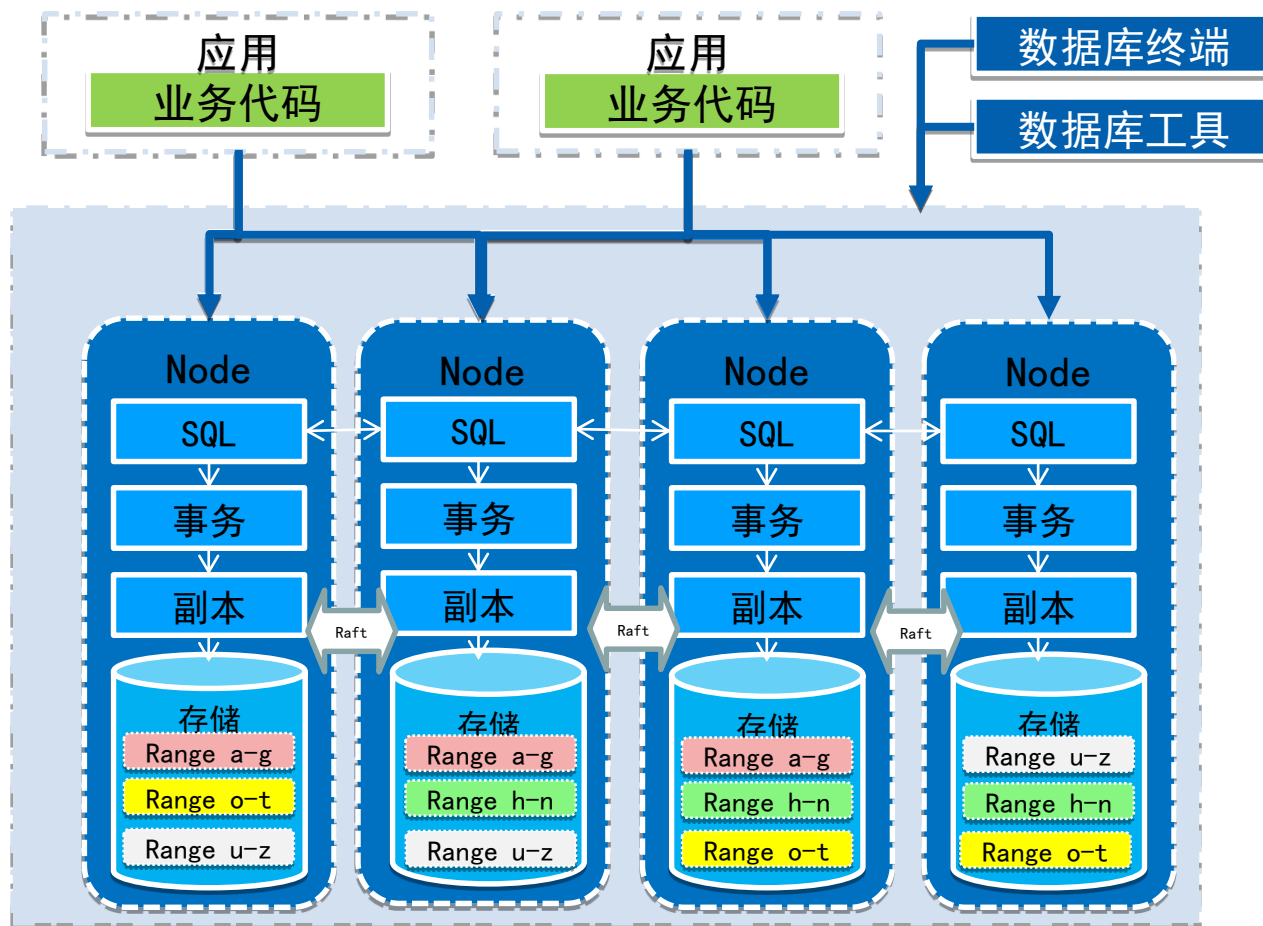
- 国产云原生数据库
- 水平扩展，支持分布式强一致事务
- 降低应用开发和运维的复杂性

一、浪潮云溪数据库简介

分布式架构

分布式架构

- ShareNothing
- 无中心节点
 - 任何节点都可以作为事务协调节点
 - 任何节点都可以作为应用连接节点
- 数据自动分Range，Range多副本
- 通过Raft协议保障数据一致性

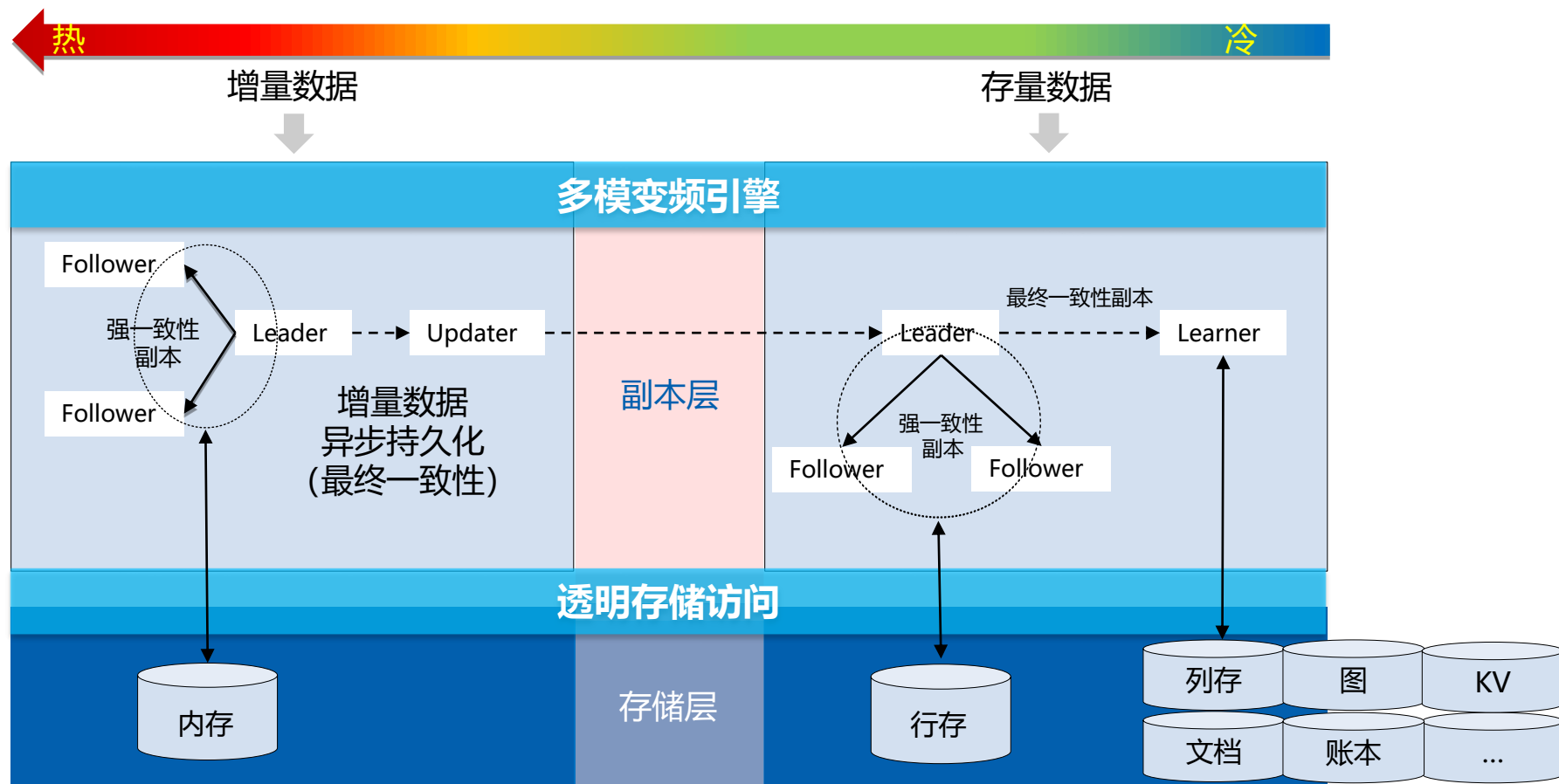


一、浪潮云溪数据库简介

多模副本

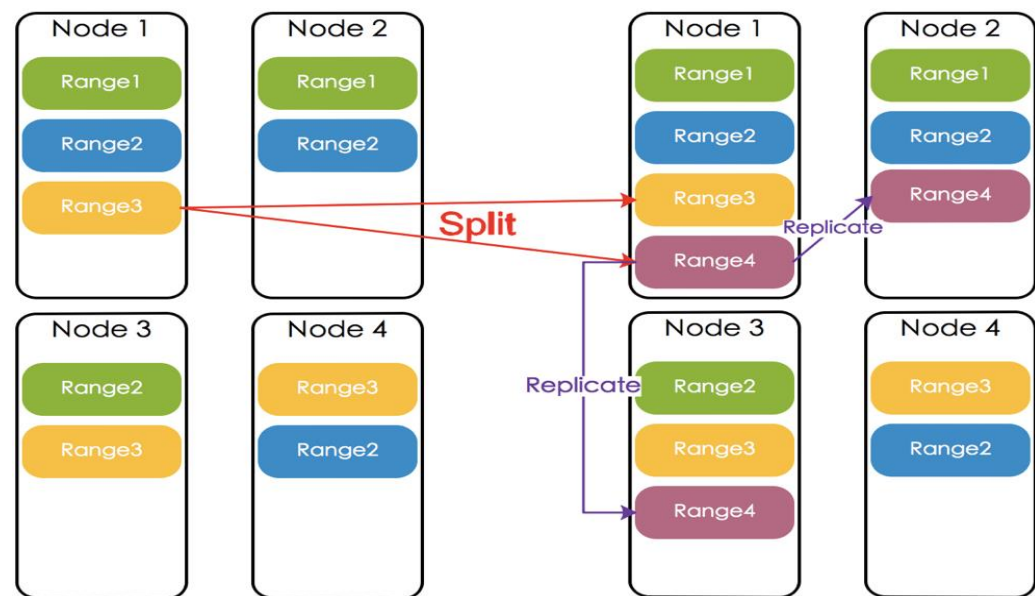
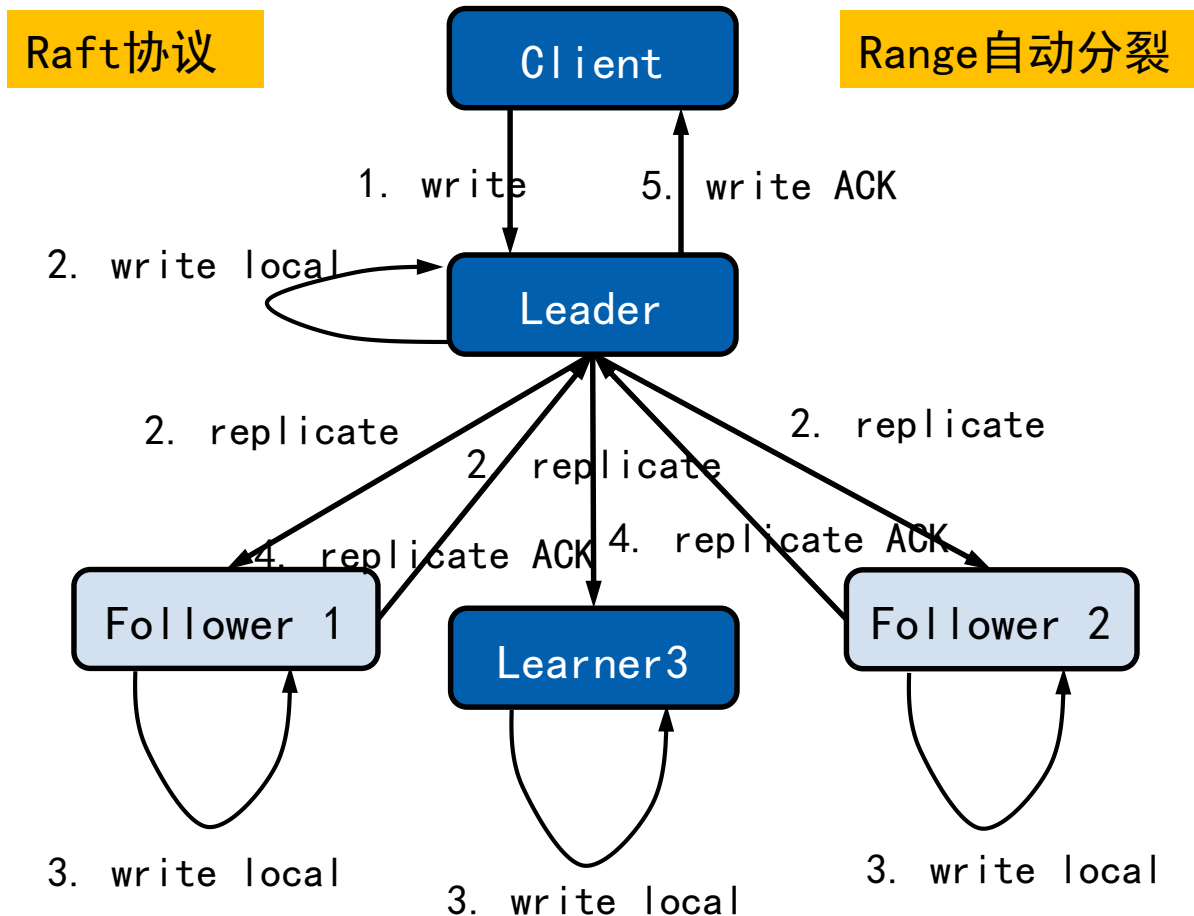
分布式架构

- 多引擎
- 透明存储
- 数据冷热分离



一、浪潮云溪数据库简介

高可用



高可用/易扩展

- 自动选举、多副本强一致、高可用
- 自动扩展

二、云溪3.0核心特性

大数据集成

需求

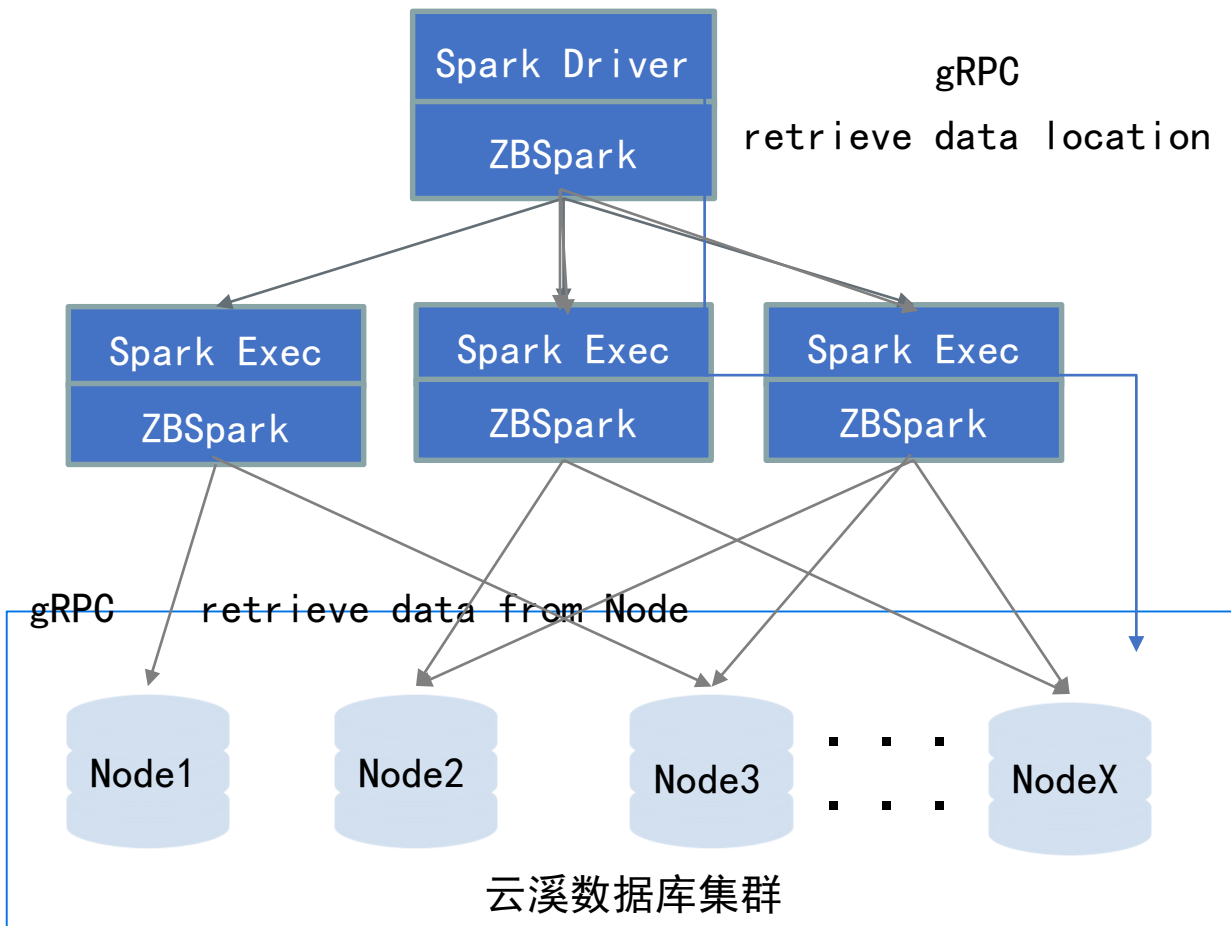
- 传统的在线ETL计算引擎为Spark，需要Spark能够从数据库高速读写，使用传统JDBC读写性能较差，不能满足需求

技术方案

- 云溪基于Spark从数据库高速读写的客户场景开发了ZBSpark插件，Spark通过RPC接口直接读取云溪存储层数据

成果

- 深度整合 Spark Catalyst，兼容多种大数据处理工具；IO提高5倍达500M/s
- 计算下推减少 Spark SQL 处理数据大小
- 支持地区亲和功能及用户权限控制



二、云溪3.0核心特性

海量数据加载

需求

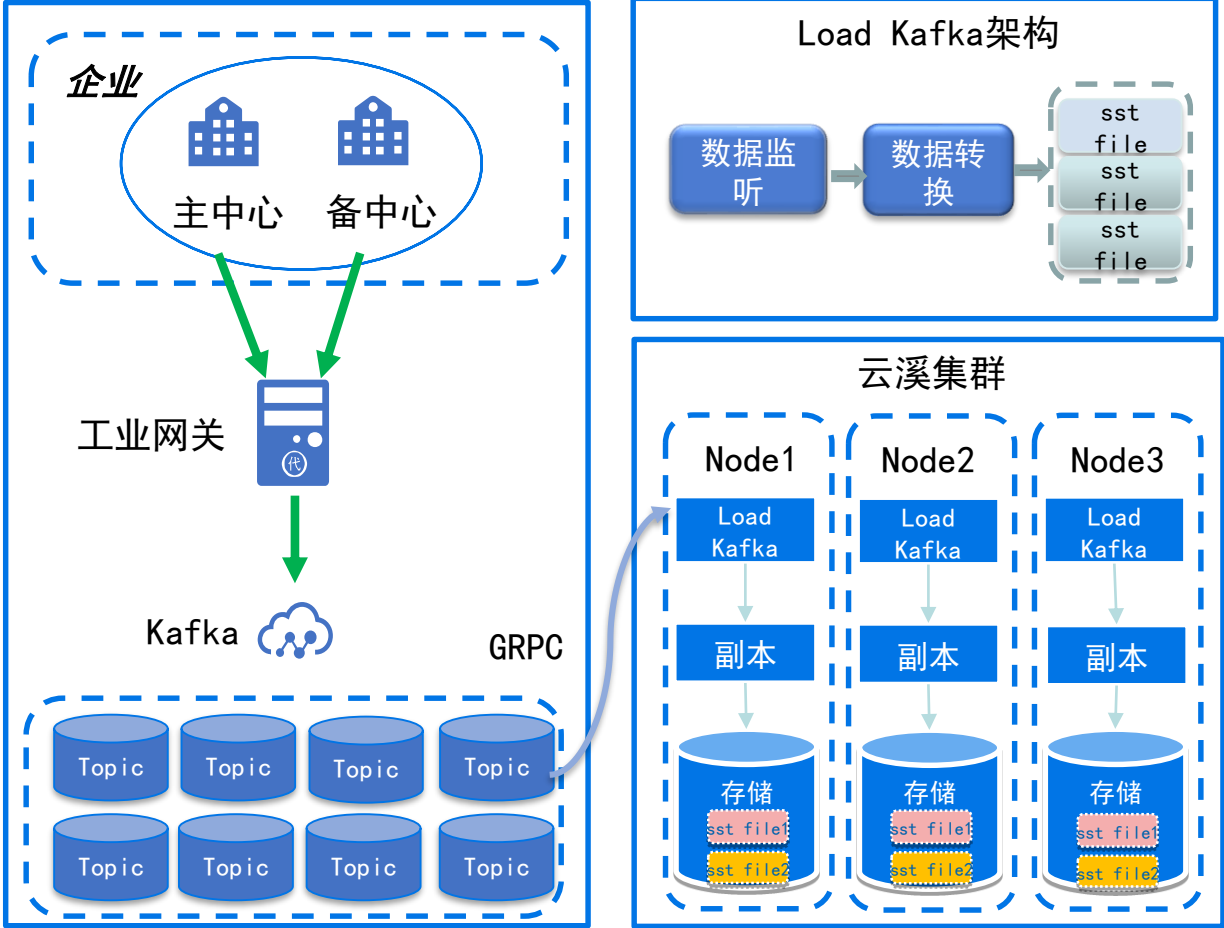
- 通过定制消息消费端，利用JDBC插入数据库，数据吞吐量上限低，性能瓶颈

技术方案

- 云开发load-kafka插件，直接从kafka读取数据
- 插件可组装导入sst文件，调研存储层接口写入数据

成果

- 原生支持kafka数据加载，导入速达度1.5G/s，数据吞吐量随节点水平增加
- 支持断点续传，从容应对加载过程中可能出现的网络波动以及硬件故障等



二、云溪3.0核心特性

HTAP

需求

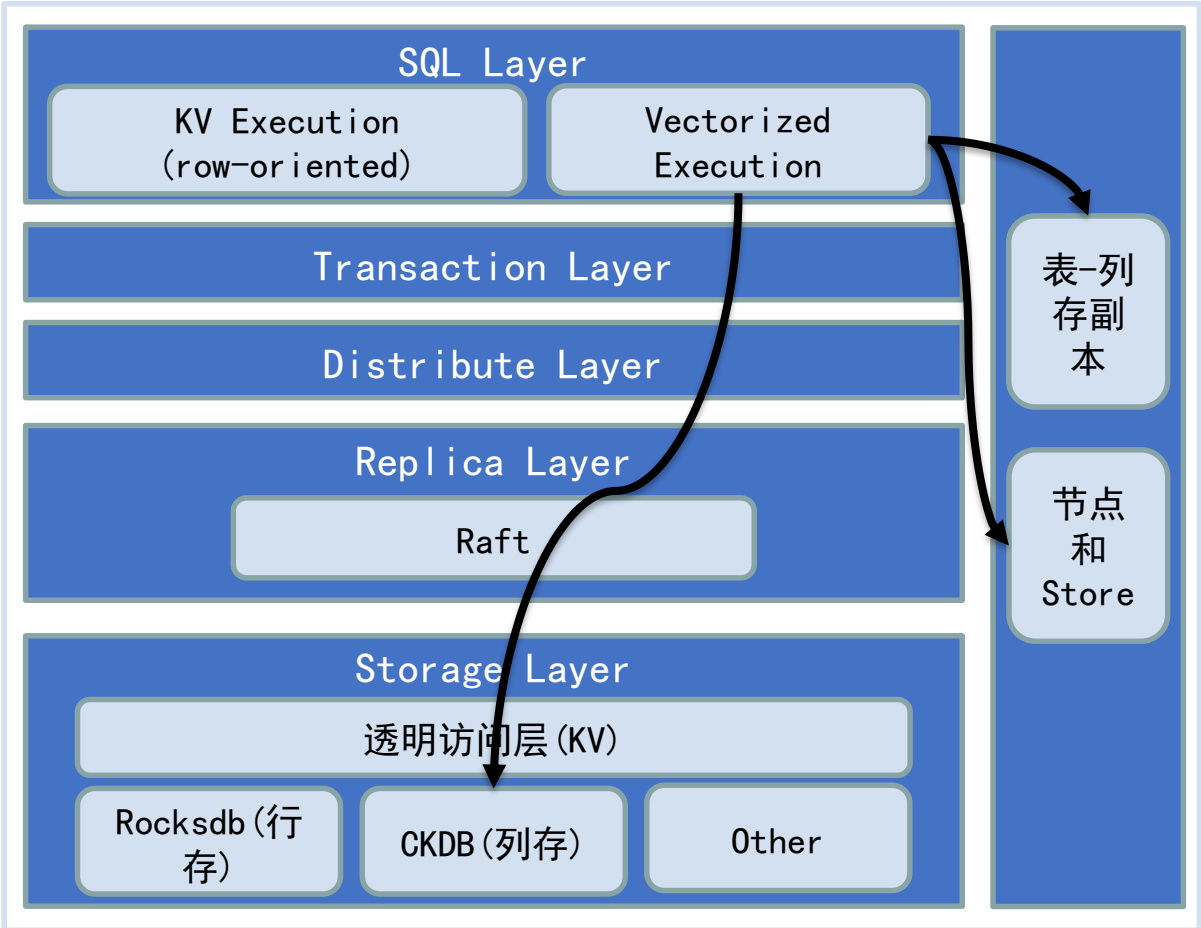
- 传统分布式关系型数据库统计分析性能较差，通过ETL建立数据仓库支持统计分析，时效差、成本高

技术方案

- 云溪数据库提供列存引擎，行列数据实时同步

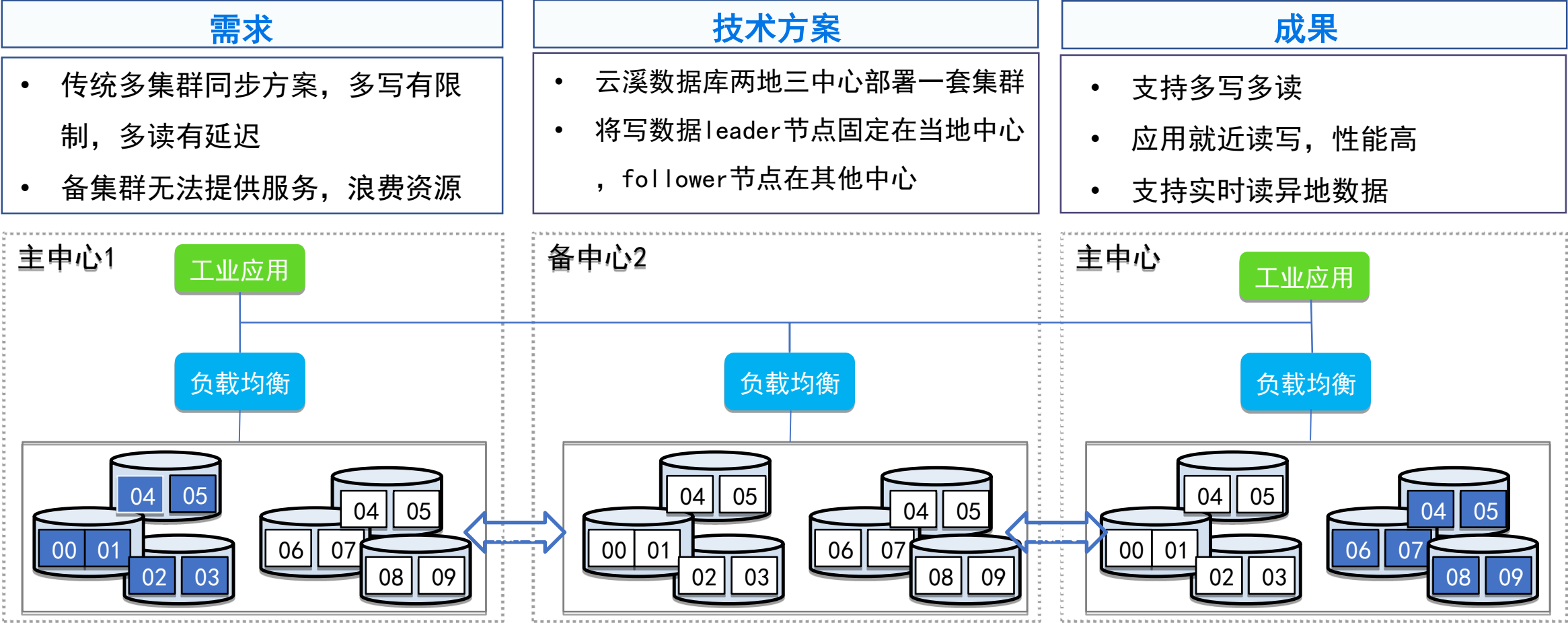
成果

- 列存副本自动同步，无需ETL，实时性高、成本低
- 计算引擎针对列存优化，包括算子下推、矢量计算，统计分析性能提高50%



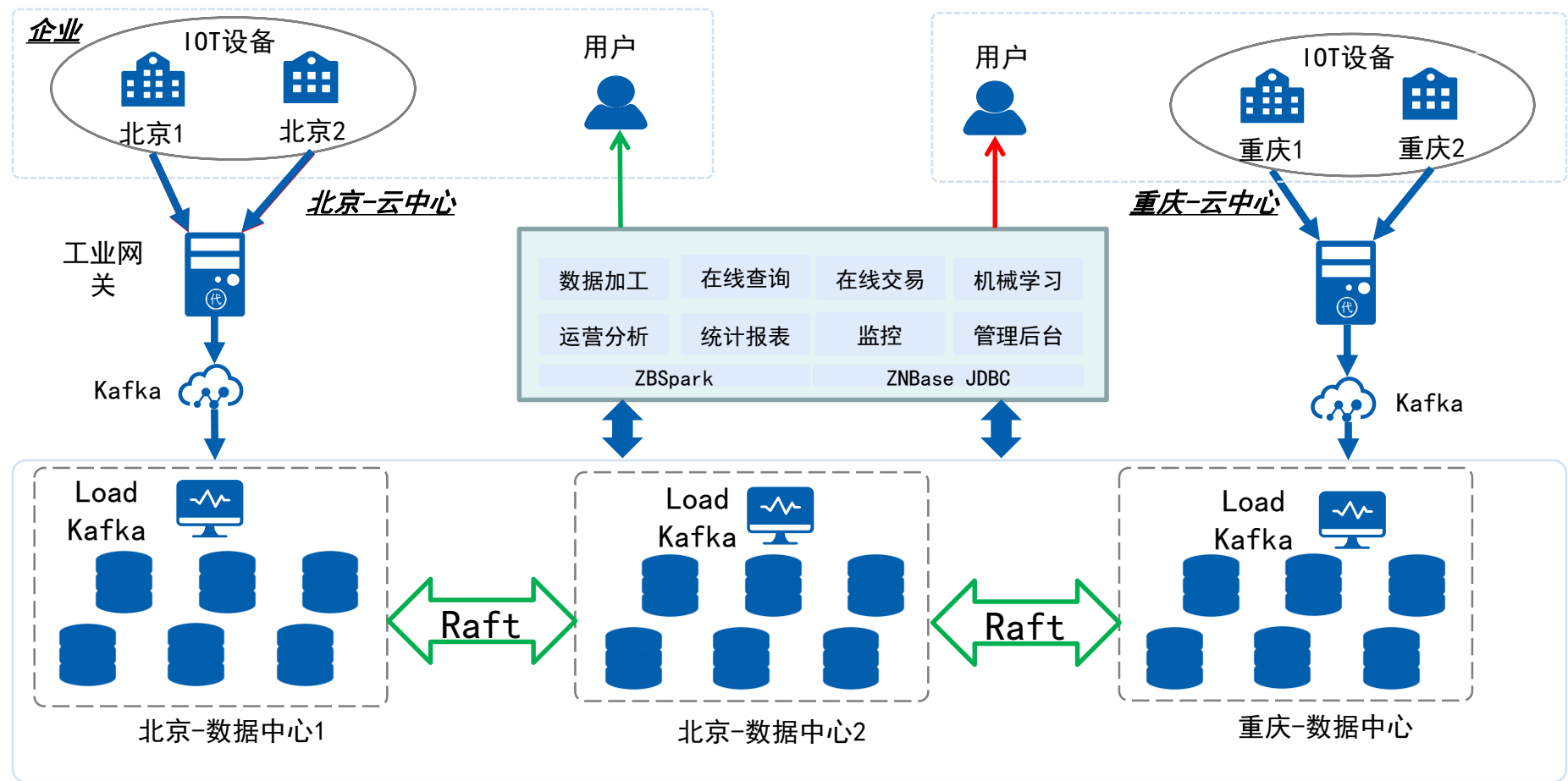
二、云溪3.0核心特性

两地三中心多活



三、应用案例

某工业大数据中心架构



云溪数据库两地三中心集群

THANKS



关注我们 / 了解更多

