# KingbaseAnalyticsDB 技术白皮书 Release V3R2

kingbase

Sep 16, 2021

# 目 录

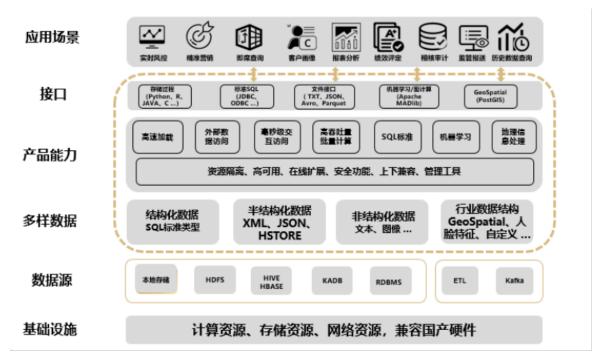
	目录	1
1	产品介绍1.1 产品简介1.2 产品特点1.3 产品架构1.4 产品兼容性1.4.1 基础设施1.4.2 操作系统1.4.3 ETL工具1.4.4 BI/报表工具1.4.5 分析工具	2 3 4 4 5 5 5 5
2	产品技术特性         2.1 高性能       2.1.1 无共享 MPP 架构         2.1.2 极速数据加载       2.1.3 高效数据存储         2.1.4 大规模并行       2.2.2 多功能性         2.2.1 全文检索       2.2.1 全文检索         2.2.2 异构数据联邦分析       2.2.3 资源配额         2.2.4 高级分析函数       2.2.5 基于 SQL 的并行计算框架         2.3 高可靠       2.3.1 主控节点可靠性         2.3.2 数据节点可靠性       2.3.2 数据节点可靠性         2.3.3 接口重试重连       2.4 并行备份与恢复         2.5 集群扩展       2.6 安全性         2.7 易用性       2.7.1 开发工具         2.7.2 管理工具       2.7.2 管理工具	66 66 66 77 99 99 100 100 111 112 122 133 144 155 155
3	容量限制	17
A	版权声明	18

# 第 1 章

# 产品介绍

### 1.1 产品简介

金仓分析型数据库系统 KingbaseAnalyticsDB(以下简称为 KADB)是北京人大金仓信息技术股份有限公司(以下简称为人大金仓)推出的 MPP 数据库产品。产品应大数据时代海量数据分析处理的需求,采用 shared-nothing 分布式架构,具有高性能,高扩展性能力,承载了人大金仓在大规模并行计算和数据库管理领域最新的研发成果,它定位于数据分析类应用市场,可以处理 TB-PB 级别的数据,并能集成多种异构数据源进行数据挖掘和分析。在处理复杂查询,如:多表连接、聚合等,也体现出了很好的性能。适用于数据仓库、决策支持、高级分析等分析类应用场景。



#### 图 1 产品全景视图

# 1.2 产品特点

KADB产品具备以下技术特点:

#### 1. LOW TCO

产品使用通用硬件,支持 X64、ARM、MIPS 等架构的物理机、云设备。无需昂贵的磁盘阵列及高性能服务器。

#### 2. 高性能数据加工

基于 MPP+shared-nothing 的分布式架构,所有数据节点能力完全对等,计算和存储能力随主机扩展线性提升,无共享架构支持可扩展至 100 台主机以上。支持列式存储、高比例压缩减少 I/O 开销,压缩等级可达到 1:20 以上,数据在压缩形态下可支持 DML 操作。

#### 3. 灵活的分布和分区方式

支持多种分布和分区方式,用户可按照自身不同的业务场景,灵活的选择相应的分布方式,数据自动根据相应的分布方式分布,无需人工干预。在处理分区表的查询时,优化器可根据查询的过滤条件,只访问这个表中有限的几个分区,减少磁盘 IO,提升查询效率。

#### 4. 毫秒级交互式查询

支持行存储,支持流水线式执行器,支持 btree、bitmap、gist、gin 等高效索引,支持固定热数据 在内存中,减少磁盘 IO 开销。在高并发的情况下,亿级别数据等值检索在毫秒级响应。

#### 5. 高可靠性

支持海量数据场景下的集群高可用特性。通过基于日志的数据同步方式,保证主控节点和数据节点主备之间的数据一致。如果相关主节点出现故障,备节点自动接替主节点继续提供服务,无需人工干预。同时,通过接口重试和重连的方式解决在主控节点切换过程中连接丢失的问题,实现业务对集群故障切换无感知的效果,当数据库集群整体不可用时,使用接口本地备份数据的方式保证故障时正在插入的数据不丢。

#### 6. 并行备份与恢复

为了缩短数据库集群系统备份和恢复耗时,KADB利用分布式的方式,实现并行逻辑备份和恢复。所有的 master 和 segment 主实例会同时对数据库对象和数据进行备份,数据库可在备份期间对外正常提供服务,备份效率可达 4GB/s。并行恢复的过程和并行备份相似,并且在恢复时会验证备份文件的合法性。集群更支持增量备份,极大的减少全量备份导致的硬件资源瓶颈问题。

#### 7. 标准符合性

支持 SQL-92 ANSI/ISO, SQL-99 ANSI/ISO, SQL 2003 ANSI/ISO 标准。支持主流的数据库接口,包括 JDBC、ODBC、C API、PYTHON API、ADO.NET 等。

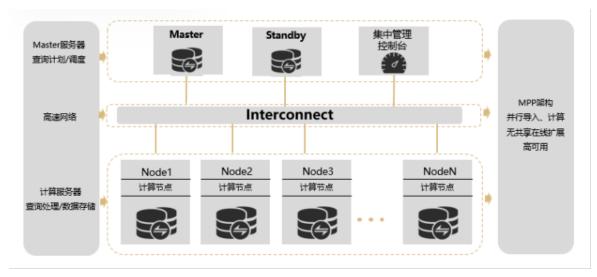
#### 8. 基于 SOL 的并行计算框架

提供基于 SQL 的并行计算能力扩展,包括 GIS 计算、机器学习、图计算和扩展自定义计算类型,学习曲线平滑。

1.2. 产品特点 3

# 1.3 产品架构

KADB 的数据库服务器部分采用 MPP+ 无共享式的分布式架构。集群组件包括集群中包括主控(Master)节点以及计算(Segment)节点。



#### 图 2 MPP 架构

主控(Master)节负责接收应用连接、为应用的请求生成执行计划并将计划分派给计算(Segment)节点执行并将执行结果汇总返回给应用。

计算(Segment)节点是独立的数据存储和查询执行引擎,用户数据以分片的方式存储在计算(Segment)节点上。计算(Segment)节点会执行主控节点分派的计算任务和将本节点的数据分发给需要数据的其他计算节点。

# 1.4 产品兼容性

#### 1.4.1 基础设施

#### 物理机

CPU: 支持 x64 指令集芯片,如 INTEL 服务器芯片、兆芯开先、海光禅定;支持 ARM 指令集芯片,如华为鲲鹏、飞腾;支持 MIPS 指令集芯片,如龙芯

网络: 支持 Ethernet 网络、Infiniband 网络

存储: 支持 HDD、SSD 存储; DRAM、AEP 内存

#### 云主机

支持主流云厂商的 IAAS 主机、云存储

虚拟机/容器

支持 VMare、docker

1.3. 产品架构 4

### 1.4.2 操作系统

支持 CentOS、RedHat 等 64 位 Linux 系统,支持中标麒麟、普华、深度等 64 位国产操作系统。

# 1.4.3 ETL 工具

支持 Kettle、Datastage、Informatica、Attunity 等工具。

# 1.4.4 BI/报表工具

支持 Cognos、Tableau、SmartBI、Fine Report 等工具。

# 1.4.5 分析工具

支持 R、SAS、SPSS、SPARK 等工具。

1.4. 产品兼容性 5

# 产品技术特性

# 2.1 高性能

### 2.1.1 无共享 MPP 架构

数据分布在所有的并行节点上,每个节点只处理其中一部分数据,所有的节点同时进行并行处理,同时由于所有节点之间完全无共享,无 I/O 冲突,所以可以做到最优化的 I/O 处理。

#### 2.1.2 极速数据加载

并行加载工具可实现文件分割异步加载、失败存盘、设置外部表分布等实用功能,并行加载工具采用独特的 Scatter/Gather 并行数据流技术,Scatter 阶段用于数据加载时,外部数据采用并行、直接方式"Scatter"到 Segment 节点,无需通过 Master 节点,因此达到极高的性能,且加载性能与 Segment 数成正比。加载速度可达 4GB/s。

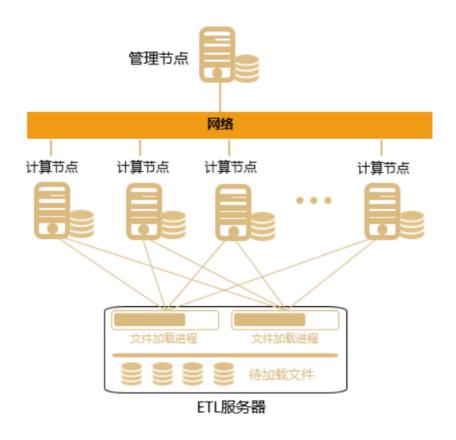
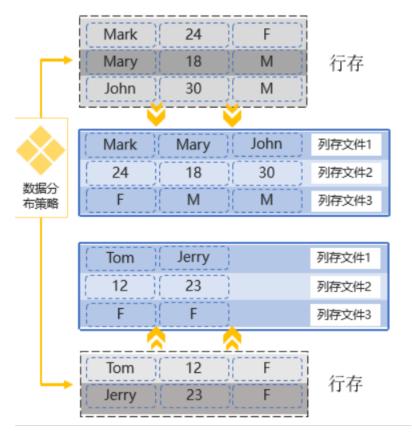


图 3 技术加载方式

#### 2.1.3 高效数据存储

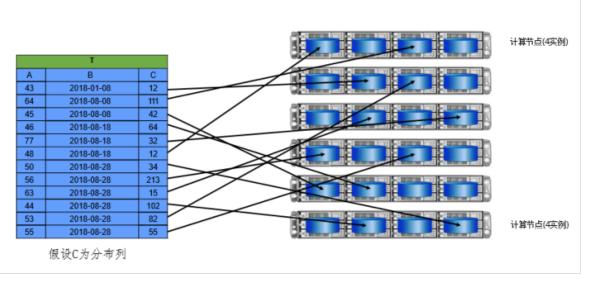
支持行存储和列存储,针对于不同的应用场景,用户可以自行选择行存表和列存表对数据进行存储,保证插入或查询性能。对于某些特定的应用场景,也可以使用行列混存的方式,将活跃的交互数据保存在行存分区表中,而非活跃的仓库数据保存在列存分区表中,空间的使用和性能的优化都能达到需求。

2.1. 高性能 7



#### 图 4 行列混存

支持多种表分布方式和分区方式,分布包括 hash 分区、随机分布和复制分布,分区包括 list、range 及混合分区,减少因数据重分布和数据扫描带来的网络以及磁盘的开销。



#### 图 5 数据分布

支持多种压缩方式,列存储支持对不同列指定不同的压缩方式和压缩等级,支持 zlib、quicklz 等高效率压缩 算法,压缩比可达 1:20;支持用户选择最适合数据的存储方式获得最佳存储效率。

2.1. 高性能 8

#### 2.1.4 大规模并行

集群支持将计算任务分发给各个主机并行执行,单个主机内多个计算节点并行执行,单个计算节点支持进一步拆分任务并行执行,可以最大程度利用集群中的计算资源完成任务。同时优化器会选择尽可能在单个主机内完成计算减少数据分发降低网络开销对计算的阻塞。

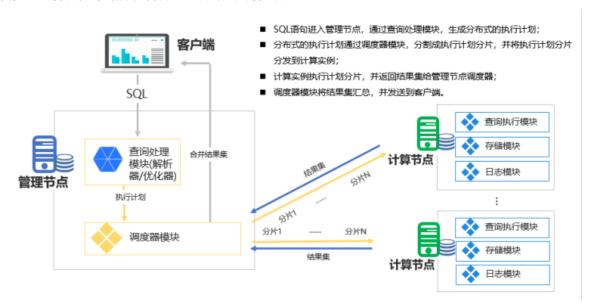


图 6 并行查询流程

# 2.2 多功能性

### 2.2.1 全文检索

支持全文检索功能,包括中文、英文、中英文混合的字词及短语的检索功能。并可以对相关需要检索的字段创建索引,加快检索效率。

2.2. 多功能性 9

#### 2.2.2 异构数据联邦分析

由于安全、数据移动开销、实时性等原因,往往很多数据不能汇聚到数据仓库,需要实时从数据源获取数据分析。KADB 支持并行的异构外部数据访问解决此类问题,支持的外部对象包括关系库如 Oracle、postgresql等;HADOOP 生态如 HBASE、HDFS 等。

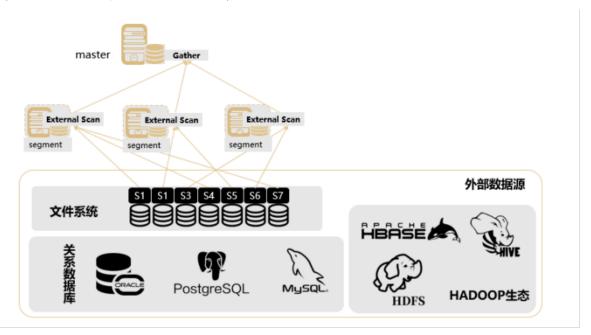


图 7 异构数据联邦

#### 2.2.3 资源配额

支持基于 CGROUP 的资源配额,控制用户级的资源使用。可用于混合应用部署避免资源争抢或实现不同优先级的计算任务。

#### 2.2.4 高级分析函数

支持高级窗口函数,用户可以使用这些高级的窗口函数用于复杂的查询处理,这些函数可以把数据在数据库内部进行计算避免了数据传输所带来的开销。

#### 2.2.5 基于 SQL 的并行计算框架

KADB 提供用户自定义函数、自定义聚集、自定义数据类型和各类主流程的 HOOK 来支持第三方基于 KADB 的分布式架构做计算能力扩展。目前支持的扩展计算能力包括:

#### **PostGIS**

用于支持地理信息数据的存储和计算。支持点、线、面相关数据类型的存储和计算,兼容 PostGIS 相关运算函数,方便应用在数据库内部实现图像数据计算操作。客户通过简单 sql 即可轻松实现以下业务场景,如"给定点方圆一公里以内的最近的 ATM 取款机","确定从目前定位行走到目标定位的路线长度","确定经过某坐标车辆 TOP-N"等。

Apache MADlib

2.2. 多功能性 10

提供图计算、机器学习、数据处理、统计分析、模型评估的常用算法。用户可在数据库中直接调用 MADlib 自带的挖掘算法创建自定义函数用于实现库内并行分析,这些算法包括了 t-统计量、p-值、朴素贝叶斯等。

#### 多语言服务器端编程

支持 R、Python、JAVA、C 等语言自定义函数,可利用语言里提供的各类工具包实现分析功能。

### 2.3 高可靠

KADB 通过副本的冗余机制来保证高可用,每个节点上的数据都会有一份冗余,在解析主节点日志后通过基于日志的数据同步方式,保证主控节点和数据节点主备之间的数据一致。如果相关主节点出现故障,备节自动接替主节点继续提供服务,无需人工干预。同时,通过接口重试和重连的方式解决在主控节点切换过程中连接丢失的问题,实现业务对集群故障切换无感知的效果,当数据库集群整体不可用时,通过接口数据本地备份的方式保证数据不丢。

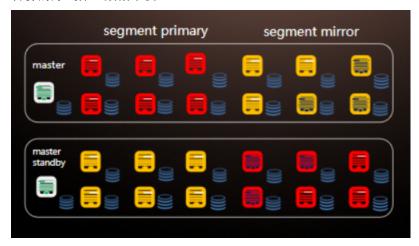
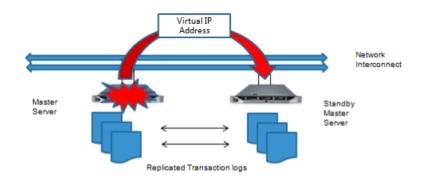


图 8 集群高可用

#### 2.3.1 主控节点可靠性

主控实例(master)和主控备实例(master standby)之间通过数据流的方式自动同步,使用虚 ip 对外提供 ip 访问接口,对外服务地址不变,在 master 实例出现故障时,脚本将 master standby 提升为 master,并设置本机虚 IP 确保对外提供服务地址不变。当故障 master 再启动时,脚本自动将其设置为 master standby 节点。整个过程无需人工手动操作。

2.3. 高可靠 11



- 1. 当Master节点failure时,将会自动切换到Standby节点。通过虚拟IP保证对外服务的地址和端口不变。
- 也可以在Standby节点上运行 dca\_failover命令将Master切换到 Standby。

图 9 master 高可用

#### 2.3.2 数据节点可靠性

KADB 中的数据都存放在 segment 节点中,如果开启了高可用特性,每一个主 segment 节点(primary)都有一个镜像 segment 节点(mirror)。主 segment 节点中的数据通过流复制的方式同步到镜像 segment 节点中,使得二者的数据保持一致。如果主 segment 发生故障,镜像 segment 节点将会自动接替主 segment 节点,继续对外提供服务。

#### 2.3.3 接口重试重连

一旦整个集群的运行环境出现对外服务中断或者机房停电的情况,实时数据通常都会被丢弃或遗漏,KADB 提供可重试和重连的接口,将可能出现影响的 sql 记录并重新发送给数据库集群,在集群上线后立马可以连 接,保证数据不丢失。当集群无法立刻上线时,接口还会将这些可能遗漏的数据保存在客户端本地的磁盘上, 在集群上线后,进行手动恢复。

2.3. 高可靠 12

# 2.4 并行备份与恢复

KADB 提供并行备份工具,让每一个节点都参与备份,提升备份效率,同时记录每个节点的备份状态,保证备份数据的原子性。同样,在恢复过程中,也会对备份文件的合理性进行校验,整个过程所有节点都会执行恢复操作,极大提升性能。

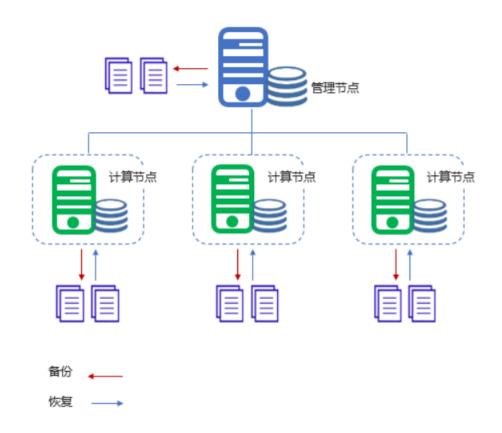


图 10 并行备份恢复

# 2.5 集群扩展

KADB采用无共享/MPP架构,可通过扩展节点来提高集群的容量和计算性能,满足用户在不同发展阶段的需求。集群支持1~数百新主机的接入,支持主机的型号和配置不同。扩展操作是动态且在线的,系统扩容在1分钟内就可完成,系统扩容后,原有数据将自动在所有节点上重分布,重分布的过程可以由用户指定在每天系统空闲的时候进行,扩展过程中无需停机,重分布数据的表支持只读,其他表在扩展中的访问无影响。

2.4. 并行备份与恢复 13

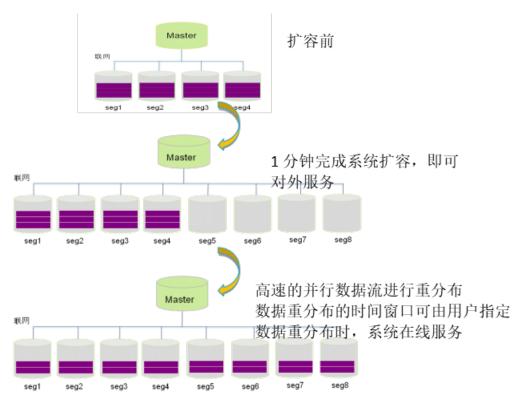


图 11 集群扩展

# 2.6 安全性

- 支持数据存储加密,支持相关加密函数,如,AES,SHA等
- 支持用户访问控制权限,最细到列级别
- 支持用户认证控制,包括密码复杂度限制、认证失败锁定、访问时间控制等行为

2.6. 安全性 14



图 12 安全功能

# 2.7 易用性

### 2.7.1 开发工具

- 支持基础对象管理,可查看表、模式、视图、数据库等对象
- 图形化执行计划显示,可通过图像显示 SQL 的执行树
- 可获取历史查询记录
- 图像化数据编辑,直接通过点击和编写交互表中的内容直接修改数据库中表的相应数据。

#### 2.7.2 管理工具

- 支持图形化部署, 无需具备专业知识, 点击下一步即可安装集群
- 可通过图形化界面部署集群高可用配置
- 利用监控工具可监控集群硬件资源使用情况,包括 CPU、内存、IO、网络等,可监控节点运行状况,对故障节点显示告警,可监控系统整体健康,保证集群正常运行

2.7. 易用性 15



图 13 监控管理工具

• 支持审计管理, 可基于库、时间、操作类型等条件进行审计, 审计结果可以文件形式导出

2.7. 易用性 16

# 第 3 章

# 容量限制

容量指	集群最大物理节点数量	1024
标	单个数据库最大容量	无限制
	单表最大容量	XXXX 无限制(单个计算节点上的单个分区为 128TB,
		32K 页面情况下)
	单表最大行数	281474976710656 (2^48)
	单表最大列数	1600
	单视图最大列数	无限制
	最大数据库数量	4,294,950,911
	单个数据库最多对象数量(表、索引、	4,294,967,296 (2^32)
	视图总量)	
	単字段最大容量	1GB
	数据库对象名的最大长度	63 个字符(支持定制增加)
	组合索引最大列数	32(支持定制增加)

# 版权声明

人大金仓版权所有,并保留对本手册及本声明的一切权利。未得到人大金仓的书面许可,任何人不得以任何方式或形式对本手册内的任何部分进行复制、摘录、备份、修改、传播、翻译成其他语言、将其全部或部分用于商业用途。

#### 免责声明

本手册内容依据现有信息制作,由于产品版本升级或其他原因,其内容有可能变更。人大金仓保留在没有任何通知或者提示的情况下对手册内容进行修改的权利。本手册仅作为使用指导,人大金仓在编写本手册时已尽力保证其内容准确可靠,但并不确保手册内容完全没有错误或遗漏,本手册中的所有信息也不构成任何明示或暗示的担保。

#### 技术支持

- 人大金仓官方网站: http://www.kingbase.com.cn/ 您可以在官网中获得人大金仓所有产品的资讯信息,销售联系方式
- 金仓数据库子网站: http://kes.kingbase.com.cn/ 您可以在产品子网站中获得最新的产品技术资料、产品 故障原因及问题分析、产品的应用解决方案、软件升级资料等等。
- 全国服务热线: 400-601-1188
- 人大金仓技术支持与反馈信箱: support@kingbase.com.cn