

# H3C DataEngine MPP 分布式数据库

## 产品概述

随着大数据技术的发展，各行各业每天都在产生海量的，多种多样格式的数据，同时也对数据库建立组织内部高性能、高效率的大数据实时分析平台提出了新的需求，以实现数亿条以上数据的秒级实时查询、更新、调用、分析挖掘等功能。传统数据库技术已经无法满足对海量数据的实时存储，计算和分析挖掘的要求，H3C DataEngine MPP数据库在此背景下不断迭代优化，整合Hadoop大数据和云技术，形成了一整套专业化的大数据解决方案。

H3C DataEngine MPP 是支撑大数据实时分析的理想平台。H3C采用Hadoop大数据和MPP分布式数据库融合的计算框架为用户提供完整的大数据平台解决方案，包括数据采集转换、存储计算、分析挖掘以及运维管理等全系列功能。H3C DataEngine MPP采用高性能的列式存储和计算技术，支持主动数据压缩，高级分析，具有弹性扩展以及自定义外部扩展等先进特性，为大数据加载、处理、导出和分析性等任务提供了无与伦比的性能、混合负载和实时分析能力，帮助用户构建海量数据处理系统，发现数据的内在价值，获取新的市场机会。

## 产品主要特性与优势

### 优化的列式存储与计算

在H3C DataEngine MPP中，每列数据是独立地存储在连续的硬盘存储块中。这与传统行式数据库按行的顺序来连续存储数据有根本的不同。H3C DataEngine MPP支持延迟物化技术。对于大多数的分析查询而言，往往只需要获取所有列数据的一个子集。H3C DataEngine MPP列式优化器和执行引擎可以在列式存储中跳过无关的列，从而节省了大量的I/O资源消耗。

### 主动压缩

与其他宣称支持列存储的数据库按照数据插入的向后顺序来存储的方式不同，H3C DataEngine MPP会根据每个列的数据类型、基数和查询特点，自动选择适用的排序方式和压缩算法，以尽可能减少数据所占的存储空间，降低查询的I/O消耗，提升查询性能。H3C DataEngine MPP支持超过12种压缩算法。

### 无共享大规模并行计算和可扩展性

H3C DataEngine MPP集群中的所有节点完全对等，不需要主节点，数据加载、数据导出和查询都可以并行地在所有节点同时执行。由于没有资源共享，增加节点就可以线性地扩展H3C DataEngine MPP的数据容量和计算能力，可以轻松从几个节点到上千节点、或从几个TB到数10PB规模扩展和收缩，满足业务规模增长的要求。

### 分级存储

在大数据时代，数据产生越来越快，而合规性和深度挖掘要求保留更多的数据，因此数据库中存放的数据越来越多。分析性能、高速磁盘高成本和大数据容量要求常常是矛盾。H3C DataEngine MPP的分级存储特性可以有效地化解这一矛盾。H3C DataEngine MPP

可以为不同的Schema、表等对象、以及表分区指定不同的存储策略，指定不同的存储位置（可以采用不同性能、成本和容量的存储介质），从而优化存储成本。

## 自动优化设计

H3C DataEngine MPP内置包含专家知识的数据库优化设计器。用户只需要指定逻辑模式（Schema），装载样例数据，并提供典型查询SQL语句，H3C DataEngine MPP的数据库优化设计器就会根据专家知识自动设计数据的水平分布方式、每个列的排序方式和压缩算法，平衡查询性能和存储空间大小要求，实现数据库整体的最优化。

## 原生云分析平台

H3C DataEngine MPP同时支持在公有云和私有云场景计算和存储资源分离，结合无共享MPP的高性能和共享存储的快速扩展优势，数据主拷贝位于共享存储易于极速扩展，热数据位于计算资源本地提供高速性能，同时通过“子集群”避免集群节点数过多在高并发负载场景下的资源浪费进而实现吞吐能力的弹性扩展。

## 产品系统架构



**无共享 MPP 架构:** H3C DataEngine MPP 采用成百上千个节点并行计算是支撑大数据分析获得足够计算和存储资源的架构关键。无共享 MPP 架构不需要共享存储，甚至不需要专门的主（首）节点，才可能扩展到成百上千个节点以支撑大数据分析。

**原生支持高可用:** H3C DataEngine MPP 内置高可用性，若在设计可用性范围内发生节点故障，可实现对应用透明，不会中断正在执行业务。H3C DataEngine MPP 的增量备份和恢复、数据库部分对象备份和恢复功能，可以在灾难发生时可以保证应用能快速恢复可用。

**优化的列式数据库:** 大数据分析性能的关键消除 I/O 开销瓶颈。H3C DataEngine MPP 列式存储和计算（列式引擎）相结合，提供包括增加、删除、修改操作在内的完整功能，是大数据条件下支持高性能分析查询和即席查询的关键。

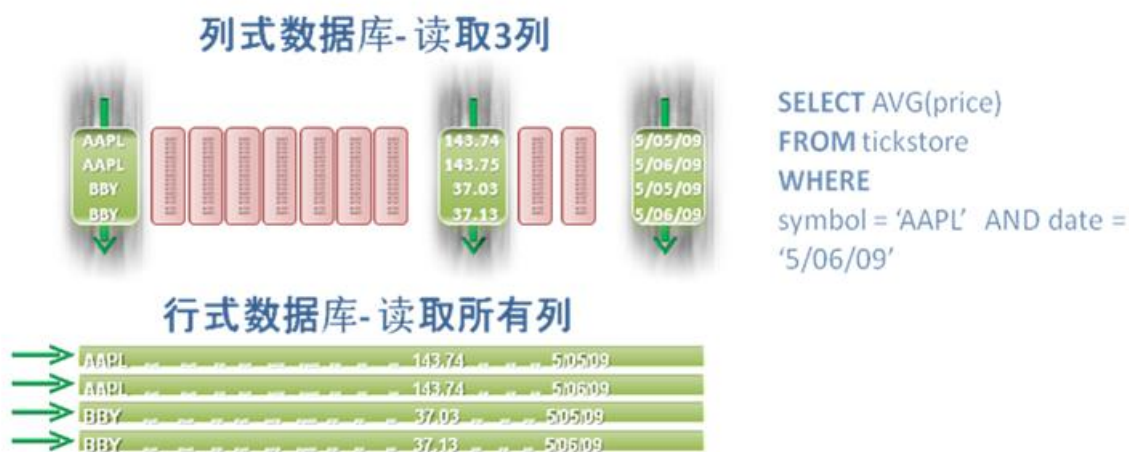
**理想的大数据实时分析平台:** H3C DataEngine MPP 采用高性能的列式存储和计算技术，支持主动数据压缩，高级分析，具有弹

性扩展以及自定义外部扩展等先进特性，是支撑大数据实时分析的理想平台。

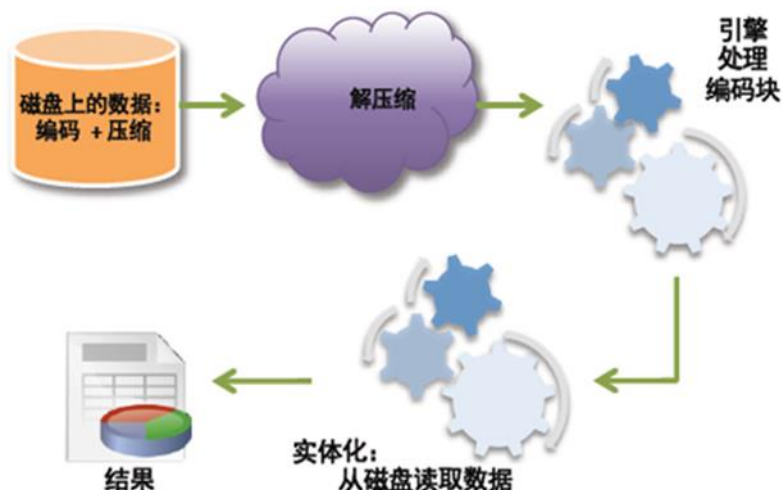
## 产品特点

### 先进的列式存储和计算

在 H3C DataEngine MPP 中，每列数据是独立地存储在连续的硬盘存储块中。这与传统行式数据库按行的顺序来连续存储数据有根本的不同。



H3C DataEngine MPP 还支持延迟解压缩技术。H3C DataEngine MPP 的优化器和执行引擎充分利用列式计算技术，支持在查询条件和关联中直接访问数据编码后的值，而不需要先解码。这样 H3C DataEngine MPP 就可以大大节省在数据查询期间的 CPU 开销，进而提升整体的查询性能。



### 高效率的主动压缩

H3C DataEngine MPP 支持超过 12 种压缩算法，如：行程长度算法 (run length encoding)，增量编码 (delta value encoding)，

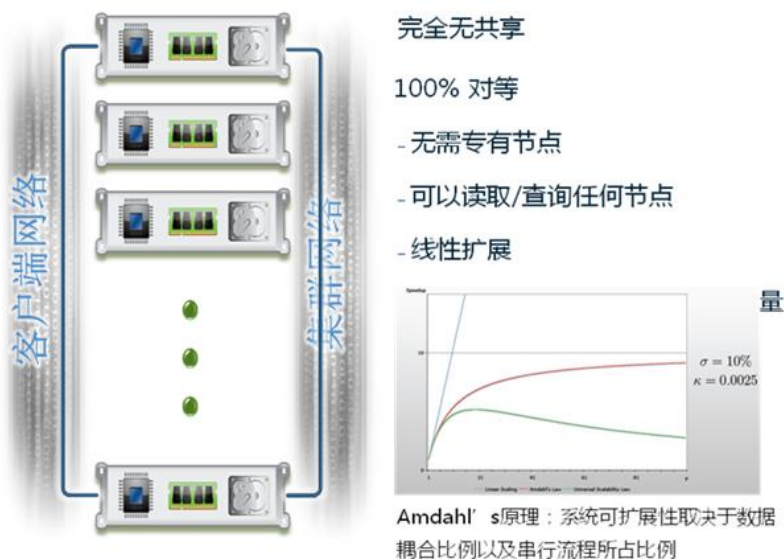
针对整数数据的整数压缩，针对字符数据的块字典编码，针对其他数据类型的 Lempel-Ziv 编码等。

H3C DataEngine MPP 的数据压缩比通常可达 10:1。从 I/O 资源消耗节约的角度来看，对 I/O 是主要瓶颈的分析系统而言，相较于传统的行式数据库，H3C DataEngine MPP 的主动压缩技术可以带来约一个数量级的性能提升。不过不同行业的业务数据，其特点差异较大。因此，H3C DataEngine MPP 对不同类型数据的压缩比也不尽相同。

## 独特便捷的无共享大规模并行计算和可扩展性

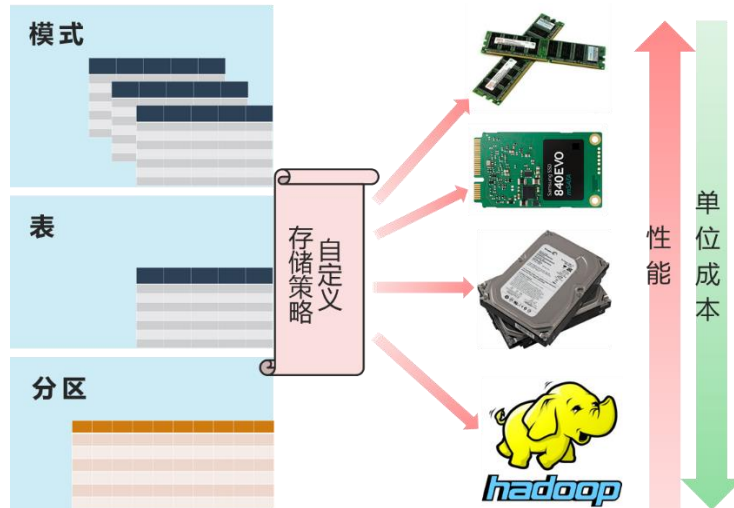
H3C DataEngine MPP 支持一键式在线群集扩展。可一键加入和删除节点并自动完成数据重分布，扩展过程不需要中断正在运行的业务。H3C DataEngine MPP 集群中的所有节点完全对等，不需要主节点，数据加载、数据导出和查询都可以并行地在所有节点同时执行。由于没有资源共享，增加节点就可以线性地扩展 H3C DataEngine MPP 的数据容量和计算能力，可以轻松从几个节点到上千节点、或从几个 TB 到数 10PB 规模扩展和收缩，满足业务规模增长的要求。

### 强大的可扩展性 -无共享的大规模并行处理(MPP)架构

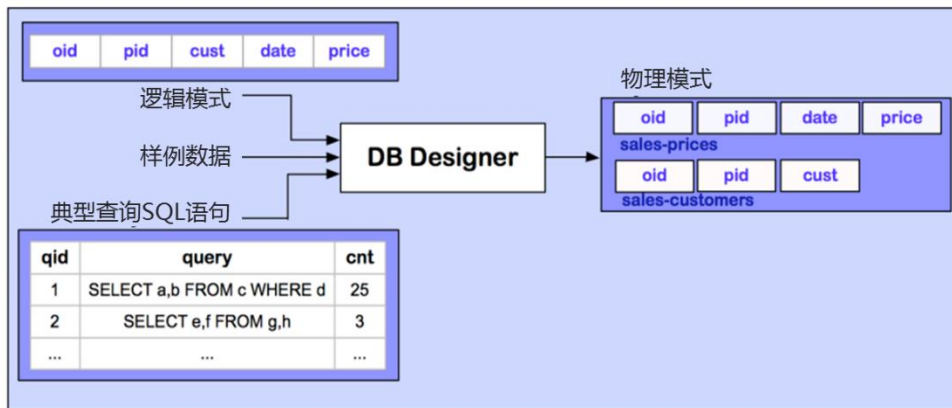


## 基于业务的分级存储

在大数据时代，数据产生越来越快，而合规性和深度挖掘要求保留更多的数据，因此数据库中存放的数据越来越多。H3C DataEngine MPP 可以为不同的 Schema、表等对象、以及表分区指定不同的存储策略，指定不同的存储位置（可以采用不同性能、成本和容量的存储介质），从而优化存储成本。同时，H3C DataEngine MPP 还支持把部分表和表分区备份到离线存储，以及从离线存储中恢复，这大大加快了数据在在线存储与离线存储间的交换速度，提升了在线设备的利用率和业务价值。



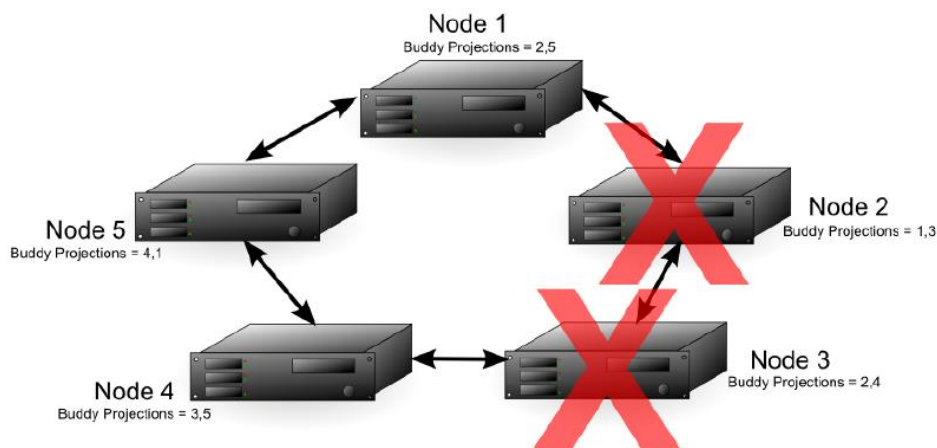
## 内置自动优化设计



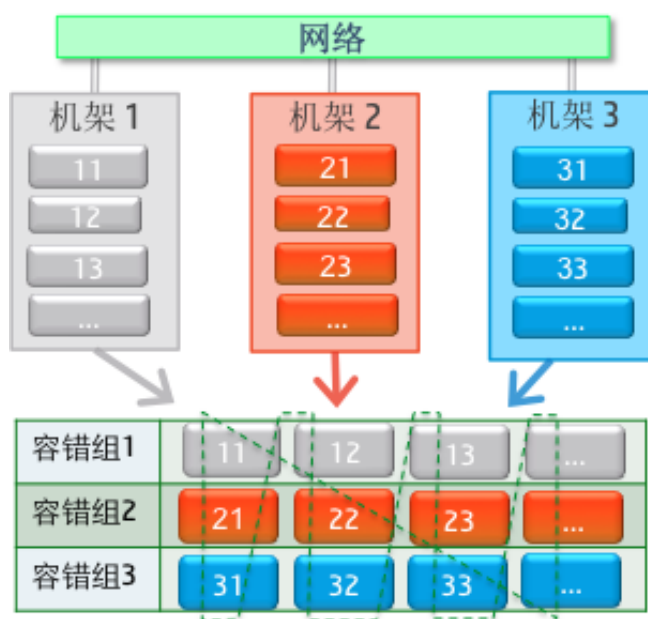
H3C DataEngine MPP 的设计器不仅提供对数据库或模式的整体优化设计，也支持针对某些特定查询的增量优化。另外，H3C DataEngine MPP 还提供负载分析器来收集数据库运行负载数据，随时提供自动化建议。从而大大降低 DBA 管理的成本。

## 稳定性保障体系

H3C DataEngine MPP 通过维护数据的多个冗余备份来实现高可用性。H3C DataEngine MPP 保证冗余数据被散列存储在不同的节点上，从而内置了智能数据镜像功能。H3C DataEngine MPP 将其称之为 K 级系数可靠性 (K-safety)，K 指的是 H3C DataEngine MPP 能够容忍的可能发生故障的任意节点的个数。



K 级系数可靠性保证了任意 K 个节点发生故障时，集群中仍然存在至少一份完整的数据来响应数据处理和查询请求。节点故障对应用透明，不会中断正在执行业务。通过调整冗余数据的份数，H3C DataEngine MPP 可以获得满足不同应用设计需要的可用性。



H3C DataEngine MPP 原生支持容错组和机架感知。H3C DataEngine MPP 的数据冗余是跨容错组的，当一个容错组（通常是一个机架，甚至是包括数个机架组成的区域）的节点都发生故障，可以保证整个数据库的数据还是完整有效的，从而有效避免机柜掉电等大规模硬件故障对整个集群可用性的伤害。

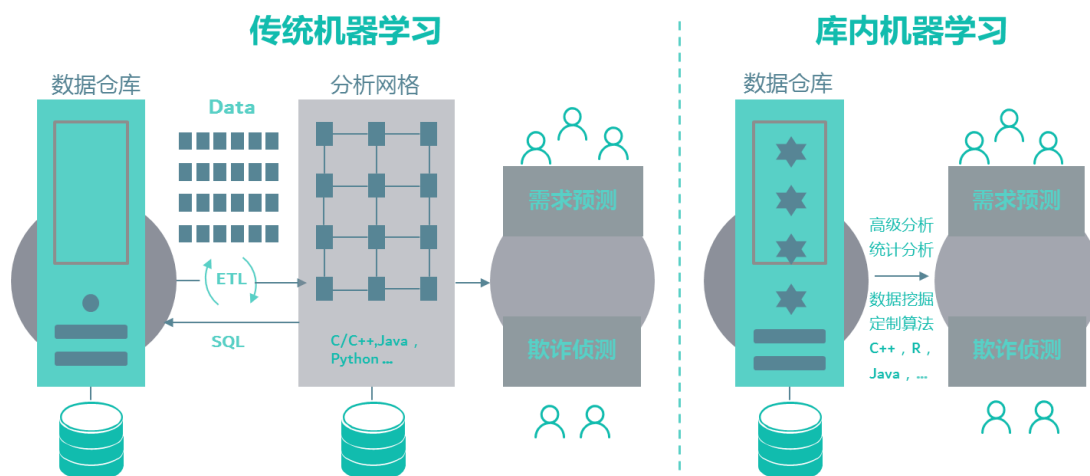
## 库内机器学习

传统数据分析系统采用“分析/挖掘请求——提起数据——分析/挖掘——展现结果”操作流程。大数据的到来，使得网络成为瓶颈，分析/挖掘任务的高延时变得让人无法忍受。

H3C DataEngine MPP 强调将分析/挖掘算法置与数据库中，让算法与数据紧密结合，充分利用 H3C DataEngine MPP 基于列式



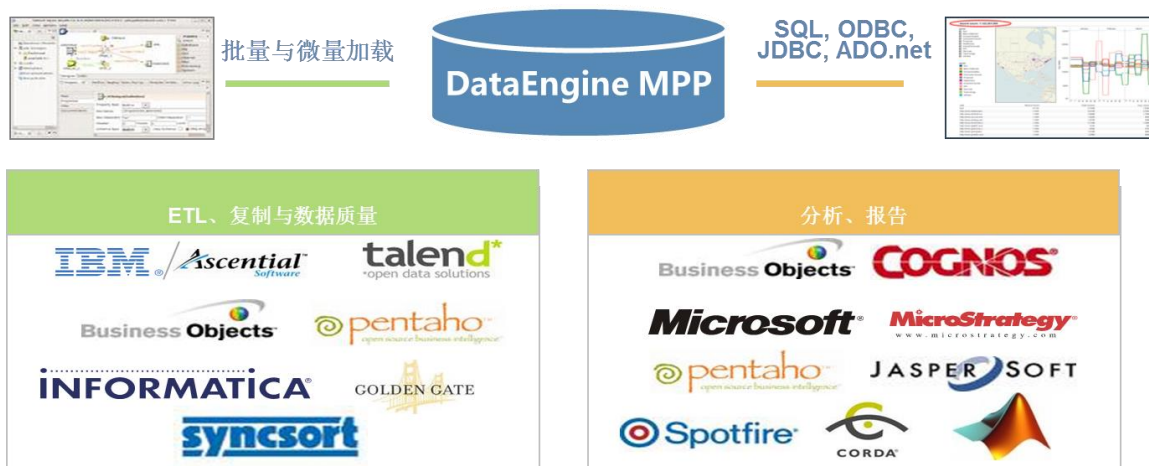
和 MPP 架构特点带来的实时计算能力，高效地执行分析/挖掘任务。



除了常见的统计分析和线性回归分析外，H3C DataEngine MPP 在 SQL-99 的基础上进行了扩展，提供了基于时间序列、模式匹配分析、地理分析等高级分析功能。这些功能将极大减少用户自己开发复杂数据分析的工作，并能很好地完成在大数据量环境下的分析任务。同时 H3C DataEngine MPP 支持结构化和非结构化融合分析，结合 H3C DataEngine MPP 与 Hadoop 分别在结构化和非结构化上的卓越分析能力，可以创造出高效的融合分析解决方案。

## 开发和扩展能力

H3C DataEngine MPP 具有很强的兼容性。支持 SQL-92/SQL-99/SQL-2003 标准，并提供 ODBC、JDBC、ADO.NET 驱动程序，可以轻松与现有的 ETL（如 Informatica、DataStage）工具、报表工具（如 BO、Cognos 等）以及分析工具（如 SAS、SPASS 等）集成，保护已有投资。

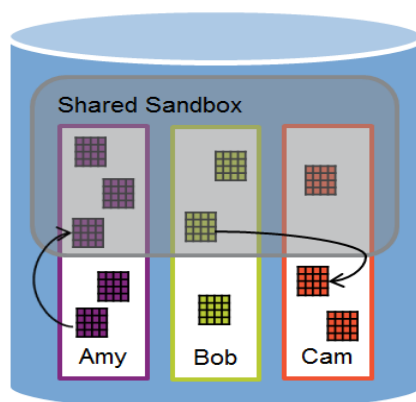


H3C DataEngine MPP 服务器采用 X86 的 Linux 服务器，支持主流的 X86 服务器和 Linux 发行版本。H3C DataEngine MPP 还提供开发自定义函数的 SDK，支持用户把 H3C DataEngine MPP 作为大数据计算平台，通过 C++/R/Java 语言来把自己独到的数据分析和处理算法作为扩展添加到 H3C DataEngine MPP 中，以充分利用 H3C DataEngine MPP 的 MPP 和列式计算技术，使用户立即拥有高效的大数据分析处理能力。

H3C DataEngine MPP 支持一键式在线群集扩展。可一键加入和删除节点并自动完成数据重分布，扩展过程不需要中断正在运行的业务。

## 数据库安全

H3C DataEngine MPP 具有很强的安全性。H3C DataEngine MPP 支持多种认证方式方式，。只有通过认证的用户才允许访问 H3C DataEngine MPP。同时，H3C DataEngine MPP 提完备的角色和授权机制，可以灵活控制用户的数据访问权限。支持多租户和沙箱，实现数据库的多租户管理模式。多租户模式可以充分利用不同租户/应用的资源消耗“峰-谷”时间差异，实现资源共享。



另外，H3C DataEngine MPP 还支持部分数据库对象（如 Schema、表）备份、恢复特性，支持按租户分别管理备份策略，使各个租户可以对自己的数据对象进行完整的生命周期管理。H3C DataEngine MPP 的数据库部分对象备份和恢复功能，为大数据系统按应用和用户的多层次分级管理提供支撑。

# H3C

### 新华三技术有限公司

北京总部  
北京市朝阳区广顺南大街 8 号院 利星行中心 1 号楼  
邮编：100102

杭州总部  
杭州市滨江区长河路 466 号  
邮编：310052  
电话：0571-86760000  
传真：0571-86760001

<http://www.h3c.com>

**客户服务热线**  
**400-810-0504**

Copyright ©2020 新华三技术有限公司保留一切权利  
免责声明：虽然 H3C 试图在本资料中提供准确的信息，但不保证资料的内容不含有技术性误差或印刷性错误，为此 H3C 对本资料中的不准确不承担任何责任。  
H3C 保留在没有通知或提示的情况下对本资料的内容进行修改的权利。