



报告名称: 数据库原理 C 课程报告

班 级: 231202

学 号: 20201000128

姓 名: 刘瑾瑾

## 浅谈数据库新技术及其发展趋势

**摘要：**数据库技术是计算机科学技术中发展最快的领域之一，也是应用最广的技术之一，已经成为计算机信息系统与智能应用系统的核心技术和重要基础。本文结合数据库技术发展的历史，通过分析数据库与其他技术结合产生的新型数据库，总结目前数据库新技术发展的一些趋势。

**关键词：**数据库，新技术，发展趋势

数据库技术起源于 20 世纪 60 年代，随着计算机技术的发展，作为计算机信息核心的数据库技术发展非常迅速，目前数据库技术已经和计算机技术紧密的结合在一起。在进 60 年的发展过程中，数据库经历了三次重大演变，造就了 C.W.Bachman、E.F.Codd 和 Jame Gray 三位图灵奖的得主。如今，数据库系统的结构已逐渐的从最开始的层次网状数据库，发展到以面向对象模型为主要特点的数据库结构。

当今数据库系统数据模型丰富多样，新技术层出不穷，应用领域也更加广泛。数据管理几乎无处不在，数据库技术和系统已经成为信息基础设施的核心技术和重要基础。对于一个国家来说，数据库的建设规模，数据库信息量的大小和使用频度已成为衡量这个国家信息化程度的重要标志。数据库已成为现代信息系统的重要组成部分，是计算机领域中发展最快的技术之一。

### 1.新一代数据库技术

新一代数据库技术近几年来在硬件的支持下得到了全面的发展，以更加丰富的数据模型和数据管理功能特征，满足更加广泛复杂的要求，发展和诞生了众多不同于第一、二代数据库系统的系统，如分布式数据库、并行数据库、图数据库和面向对象数据库等。

#### 1.1 分布式数据库

分布式数据库系统是目前数据库技术发展的主要的系统之一，其主要由一个网络不同的计算机组合在一起，数据在物理上分布而在逻辑上集中。与普通系统的不同，这种系统每一台服务器上都有独立的数据库系统和计算机相关的硬件和软件。在这些系统中可以利用局域网的形式把所有的服务器连接起来，而且这些连接起来的服务器可以独立的处理相关的数据，互不影响。

分布式数据库的核心技术包括数据复制，即不同副本之间的数据同步机制；数据分区，即如何将海量数据分布到不同节点中；分布式事务，即解决多节点面临的原子性、一致性、隔离性、持久性挑战。

#### 1.2 并行式数据库

并行处理技术与数据库技术的结合，具有潜在的可行性。因为关系数据库模型本身就有极大的并行可能性。关系数据库模型中，数据库是元组的集合，数据库操作实际是集合操作，许多情况下可分解为一系列对子集的操作，许多子操作不具有数据相关性，因而具有潜在的并行性。并行数据库系统是在并行机上运行的具有并行处理能力的数据库系统。并行数据库系统是数据库技术与并行计算技术相结合的产物。

从硬件结构来看,根据处理机与磁盘及内存的相互关系可以将并行计算机分为三种基本的体系结构:共享内存结构、共享磁盘结构和无共享资源结构。SM结构又可称为完全共享型,所有处理机存取一公共全局内存和所有磁盘;SD结构中每个处理机有自己的私有内存,但能访问所有磁盘;SN结构中所有磁盘和内存分散给各处理机,每个处理机只能直接访问其私有内存和磁盘,各自都是一个独立的整体,处理机间由一公共互连网络连接。

### 1.3 面向对象数据库系统

面向对象数据库系统是数据库技术与面向对象程序设计方法相结合的产物。面向对象数据库系统将数据作为能自动重新得到和共享的对象存储,包含在对象中的是完成每一项数据库事务处理指令,这些对象可能包含的数据不是一个类型,包含传统的数据与处理过程,也包含视频、图像和声音信号。

面向对象的数据库系统通过建立新的多媒体应用能力使得软件开发变得简单,不同类型的数据由此联系结合起来。面向对象的数据库技术是一种比较新的新的技术,相关理论知识还不完善,在处理很多事务的数据方面和关系数据库系统相比慢的多。但是,研究者们已经开发出了混合关系对象数据库,由此将面向对象数据库系统处理新型数据复杂关系的能力与关系数据库管理系统处理事务的能力很好的结合起来。

### 1.4 主动数据库系统

主动数据库是相对于传统数据库的被动性而言,它在后者的基础之上,结合面向对象技术和人工智能技术而产生,它能够根据环境的变化或是各类事件的发生主动提供用户相应的信息服务。传统数据库系统之所以是被动的,是因为它只能根据应用程序或是用户的服务请求存储或是检索数据库,不能主动根据数据库的状态或是发生的事件做出反应。主动数据库就比较灵敏,它将事件—条件—动作这一传统规则嵌入到传统数据库系统中,在紧急情况下就能及时地对事件做出反应。相当于系统提供了一个“自动监测”机构,主动不时检查这些规则中包含的事件是否已发生,一旦事件被发现,就主动触发执行相应的动作。

主动数据库主要解决的问题有知识模型、执行模型、条件检测、事务调度、体系结构和系统效率。系统实现的关键技术是条件检测,条件可以是动态条件、多重条件、交叉条件。目前已取得的成果有:多重条件同时求值算法;求值过程中间结果的生成和维护方法;递增求值方法;求值时利用规则动作部分知识;代价模型和启发式方法。

### 1.5 模糊数据库系统

现实世界中,确切的值只是特例,而模糊的值却更一般、更自然。研究模糊数据和模糊数据库有着重要的理论价值和现实意义。模糊数据库中,数据的模糊性是模糊数据库的第一公共属性,数据间的联系的模糊性为模糊数据库的第二属性,同样在约束条件、数据操作、查询语言、用户视图与外模式、数据冗余、数据间函数相关性等方面都具有模糊性。模糊数据模型是研究模糊数据库的基础,模糊数据模型有模糊关系数据模型、模糊网状数据模型,模糊层次数据模型,模糊 E-R 模型,面向对象的模糊数据模型、模糊逻辑模型和模糊知识模型等。

由于模糊数据模型和模糊数据库语言比精确时要复杂得多,因此在实现模糊数据库管理系统时,无论功能上、逻辑结构上、物理组织上、存储形式上都相应

复杂得多。模糊数据库的理论和技术的研究仍处于发展阶段，但在模式识别、过程控制、案情侦破、医疗诊断、工程设计、专家系统等领域得到了较好的应用，具有广阔的前景。

### 1.6 图数据库

图数据通过将实体与关系点变化的方式将知识结构化保存，是一种基于事务关联关系的模型表达，具有数据天然可解释性，备受学术界和工业界推崇。在数据的关联分析中，传统的关系型数据库需要进行大量的关联操作，在小规模数据的情况下这样的操作还可以接受，但是当数据规模逐渐增大，关联操作会造成性能呈指数级下降。图数据库相较于传统关系型数据库和 NoSQL 数据库，其丰富完整的关系表达提供了高效的关联查询和完备的实体信息。

图是一组对象的集合，由顶点和边构成，顶点表示实体或实例，如人、账号、组织、业务等，它们可以类比于关系型数据库里的记录或行，或文档数据库里的文档；边是连接顶点的线，表示顶点之间的关联关系。边是图数据库中的关键概念，在关系型数据库或文档存储数据库中，没有对边的抽象概念进行直接实现。图模型主要包含属性图、资源描述框架（RDF）。

图数据库基于图模型，对图数据进行存储、操作和访问，与关系型数据库中的联机事务处理（OLTP）数据库是类似的，支持事务、可持久化等特性。图数据库根据底层存储实现的不同，可分为原生图数据库和非原生图数据库两种。原生图数据库：使用图模型进行数据存储，可以针对图数据做优化，从而带来更好的性能；非原生图数据库：底层存储使用非图模型进行存储，在存储之上封装图的语义，进行图处理，其优点是易于开发，适合产品众多的大型公司，形成相互配合的产品栈。

### 1.7 多媒体数据库

现如今，多媒体技术的发展不断深入，多媒体应用与我们的生活已经密不可分，在多媒体应用之中必然涉及到海量的多媒体信息，包括视频、声音、图像、文本、图形等信息。海量的多媒体信息的应用与共享，必然需要数据库的支撑，由此多媒体数据库技术作为数据库技术的一个分支发展起来，并由此建立多媒体信息系统。

多媒体数据库应具备的功能如下：能表示和理解多媒体数据，能刻画、管理和表现各种媒体数据的特性和相互关系；具备物理数据独立性和媒体数据独立性，媒体类型可扩展；提供更为灵活的模式定义和修改功能，支持模式进化与演变，具备某些长事务处理的能力；提供多媒体访问的各种手段，近似性查询，混合方式访问等。

### 1.8 Web 数据库

数据库技术与 Web 技术相结合，出现了 Web 数据库。Web 数据库由数据库与 Web 技术结合产生，数据库服务器的工作是将数据集中起来并快速进行检索，它在后台运行。接下来的任务就交由 Web 服务器来完成，Web 服务器主要负责解决用户的需求，接收到用户的要查询的内容之后，Web 服务器会通过 Web 页面的应用程序或是脚本，根据用户的查询内容查找相关数据并将查询结果回传至客户端，之后用户就能在客户端的浏览器看到查询的结果。Web 服务器介于客户端浏览器（前台）和数据库服务器（后台）两者之间，相当于一座桥梁，让用

户能顺利通往 Web 数据库查找需要的资源信息。数据库技术的发展要比 Web 技术成熟，将前者用于组织和管理数量较大的数据十分的方便；当然 Web 技术也有数据库技术无法媲美的优点，它不仅简单易学，与平台无关这一特点使安装维护和用户培训等费用大大降低。

## 2. 数据库新技术发展趋势

大数据时代，数据量不断爆炸式增长，数据存储结构也越来越灵活多样，日益变革的新兴业务需求催生数据库及应用系统的存在形式愈发丰富，这些变化均对数据库的各类能力不断提出挑战，推动数据库技术不断向着模型拓展、架构解耦的方向演进，与信息集成、嵌入式、传感器技术、区块链、云计算、人工智能、新型硬件等技术呈现取长补短、不断融合的发展态势。

### 2.1 Web 信息集成技术

随着 Internet 和 Web 技术的迅速发展，网络已成为一种重要的信息传播和交换的手段，尤其是在 Web 上有着丰富的数据来源。如果能把 Web 上的有用数据利用起来，Web 将成为一个全球性的信息集成系统，由全世界共享。信息集成系统的方法有两种：数据仓库方法和 Wrapper/Mediator 方法。

数据仓库法按照需要全局模式从各数据源抽取并转换数据，存储在数据仓库中，用户的查询是对数据仓库中的数据进行查询。

另一种方法是 Wrapper/Mediator 方法。该方法并不将各数据源的数据集中存放，而是通过 Wrapper/Mediator 结构满足上层集成应用的需求。这种方法的核心是中介模式。信息集成系统通过中介模式将各数据源的数据集成起来，而数据仍存储在局部数据源中，通过各数据源的包装器对数据进行转换使之符合中介模式。用户的查询基于中介模式，不必知道每个数据源的特点，中介器将基于中介模式的查询转换为基于各局部数据源的模式查询，它的查询执行引擎再通过各数据源的包装器将结果抽取出来，最后由中介器将结果集成并返回给用户。Wrapper/Mediator 方法解决了数据的更新问题，从而弥补了数据仓库方法的不足。

影响 Web 数据源集成的主要问题是异构。异构表现在三个方面：（1）模式异构。模式异构表现在不同的数据源具有不同的存在形式，有的以数据库的形式存在，有的以文件形式存在，有的以 HTML 的形式存在。即使某些数据源以数据库的形式存在，它们的数据模型、存储结构及其所依赖的数据库管理系统也可能大不一样。具有相同数据模型的数据源所依赖的数据库管理系统也不尽相同；（2）数据异构。数据异构表现在不同的数据源具有不同的数据类型，有的数据源可能仅包含数字和字符之类的简单数据，有的则可能包含多媒体数据，有的可能包含时间序列等复杂数据；（3）语义异构。语义异构表现在相同的数据形式表示不同的语义或同一个语义由不同形式的数据表示。

Web 信息集成研究还有另外两个挑战：第一个挑战是如何模型化源数据内容和用户查询；另一个挑战是当数据源的查询能力受限时，如何处理查询和进行优化。

### 2.2 嵌入式移动数据库技术

研究移动计算环境中的数据管理技术，已成为目前分布式数据库研究的一个新的方向，即移动数据库技术。与基于固定网络的传统分布计算环境相比，移动

计算环境具有以下特点：移动性、频繁断接性、带宽多样性、网络通信的非对称性、移动计算机的电源能力、可靠性要求较低和可伸缩性等。

移动计算以及它所具有的独特特点,对分布式数据库技术和客户/服务器数据库技术,提出了新的要求和挑战。移动数据库系统要求支持移动用户在多种网络条件下都能够有效地访问所需数据,完成数据查询和事务处理。通过移动数据库的复制/缓存技术或者数据广播技术,移动用户即使在断接的情况下也可以继续访问所需的数据,从而继续自己的工作,这使得移动数据库系统具有高度的可用性。此外,移动数据库系统能够尽可能地提高无线网络中数据访问的效率和性能。

移动数据库系统通常应用在掌上电脑、PDA、车载设备、移动电话等嵌入式设备中,又称嵌入式移动数据库系统。嵌入式移动数据库涉及数据库技术、分布式计算技术、移动通信技术等多个学科领域。为保证数据的安全性,采取的主要措施有:(1)对移动终端进行认证,防止非法终端的欺骗性接入;(2)对无线通信进行加密,防止数据信息泄漏;(3)对下载的数据副本加密存储,以防移动终端物理丢失后的数据泄密。

移动数据库管理系统是一种动态分布式数据库管理系统,嵌入式移动数据库管理系统在移动计算的环境下应用在嵌入式操作系统之上,它必须考虑的设计原则有:(1)微内核结构;(2)对标准 SQL 的支持;(3)自动维护事务的完整性、原子性等特性,支持实体完整性和引用完整性;(4)具有完善的数据同步机制;(5)支持串行通信、TCP/IP、红外传输、蓝牙等多种连接协议;(6)具有自动恢复功能,保证用户数据的安全可靠;(7)支持 Windows CE、Palm Os 等多种嵌入式操作系统。另外,设计时还应考虑移动设备的计算能力小、存储资源少、带宽有限及 Flash 存储上写操作速度慢等特性。

### 2.3 微小型数据库技术

随着移动计算时代的到来,嵌入式操作系统对微小型数据库系统的需求为数据库技术开辟了新的发展空间。微小型数据库技术目前已经从研究领域逐步走向应用领域。一般说来,微小型数据库系统(a small-footprint DBMS)可以定义为:一个只需很小的内存来支持的数据库系统内核。微小型数据库系统针对便携式设备其占用的内存空间大约为 2MB,而对于掌上设备和其他手持设备,它占用的内存空间只有 50KB 左右。内存限制是决定微小型数据库系统特征的重要因素。微小型数据库系统根据占用内存的大小又可以进一步分为:超微 DBMS、微小 DBMS 和嵌入式 DBMS 3 种。

微小型数据库系统与操作系统和具体应用集成在一起,运行在各种智能型嵌入设备或移动设备上。微小型数据库技术目前已经从研究领域向广泛的应用领域发展,各种微小型数据库产品纷纷涌现。尤其是对移动数据处理和管理需求的不断提高,紧密结合各种智能设备的嵌入式移动数据库技术已经得到了学术界、工业界、军事领域和民用部门等各方面的重视并不断实用化。

### 2.4 传感器数据库技术

随着微电子技术的发展,传感器的应用越来越广泛。根据传感器在一定的范围内发回的数据,在一定的范围内收集有用的信息,并且将其发回到指挥中心。当有多个传感器在一定的范围内工作时,就组成了传感器网络。传感器网络由携带者所捆绑的传感器及接收和处理传感器发回数据的服务器所组成。传感器网络

中的通信方式可以是无线通信，也可以是有线通信。

传感器节点上数据的存储和处理方法有两种：（1）传感器数据存储在一个节点的传感器堆栈中，这样的节点必须具有很强的处理能力和较大的缓冲空间；（2）适用于一个芯片上的传感器网络，传感器节点的处理能力和缓冲空间是受限制的，在产生数据项的同时就对其进行处理以节省空间，在传感器节点上没有复杂的处理过程，传感器节点上不存储历史数据；对于处理能力介于（1）和（2）传感器网络的网络来说，则采用折衷的方案，将传感器数据分层地放在各层的传感器堆栈中进行处理。

传感器网络越来越多地应用于对很多新应用的监测和监控。新的传感器数据库系统需要考虑大量的传感器设备的存在，以及它们的移动和分散性。因此，新的传感器数据库系统需要解决一些新的问题。主要包括：传感器数据的表示和传感器查询的表示、在传感器节点上处理查询分片、分布查询分片、适应网络条件的改变、传感器数据库系统等。

## 2.5 区块链数据库技术

传统数据库模式存在很多缺陷：（1）不同的数据库可能存储着相同的用户基本身份信息，导致数据冗余度高；（2）不同的中心机构各自管理自己的数据，不利于机构之间的数据共享；（3）每个数据库大都由单一机构中心化管理，使得用户必须无条件信任该机构，存在中心化问题；（4）用户不能够独立验证数据的正确性，如果数据被恶意篡改，用户与机构都无法察觉。

区块链具有去中心化、信息不可篡改等特征，区块链数据库能够长期留存有效记录，保护数据不被篡改，数据库的所有历史操作均不可更改并能追溯，为解决上述这些问题提供了可能。

为此，我们提出了区块链数据库的概念，核心思想是：通过限制中心机构对数据记录的操作，来达到防篡改和去中心化的目的。该数据库中有多条区块链，每一条区块链相当于传统数据库中的一张表，所有的中心机构充当数据的存储节点，所有的存储节点根据共识算法生成区块链，所有节点(包括用户)存储区块头信息，可以由区块头信息检索到记录并验证记录的正确性。我们希望有高效的共识算法来提高系统的吞吐率,并有高效的查询算法实现在区块链上检索数据。当前，在共识算法上已经有很多研究，例如 POW, POS, PBFT。

区块链数据库典型产品有 Blockchain DB、Bigchain DB 和 Chain SQL 等。该类产品具体研究问题主要分为数据存储与事物处理两大类，存储方向分为键值对和关系型数据存储，事物处理方向聚焦于在区块链上完成数据库的事物并发控制、访问控制授权、查询处理优化等传统问题。

区块链数据库由于要容忍节点拜占庭行为而不得不采用代价更高的 PBFT、POW 等共识算法成为落地应用的一大挑战，此外，由于没有统一的协调者，如何保证区块链网络分片时分布式系统的安全性，高并发下的并行控制如何保证 ACID 也都是设计者不可忽视的问题。

## 2.6 多模数据库技术

后关系型数据库阶段，数据结构越来越灵活多样，如表格类型的关系数据、半结构化的用户画像数据以及非结构化的图片和视频数据等。多模态数据建模将

由目前以关系为中心逐步转为以对象为中心,更多关注非结构化数据的语义建模,强调数据与操作的可封装性、多重分类和动态分类,尤其是多模态数据之间复杂的语义关联,未来不是模式限定数据变化,而是数据驱动模式演化。

多模数据库,是一个集成式智能型数据库,能够执行存储、访问、处理不同类型数据等各种任务。多模数据库支持灵活的数据存储类型,将各种类型的数据进行集中存储、查询和处理,可以同时满足应用程序对于结构化、半结构化和非结构化数据的统一管理需求。

多模数据库包含 Arango DB, Orient DB, Marklogic Server, Mongo DB 等多个数据库,可以处理包括文档、图片或键值、RDF、XML 等模型,具有执行监视或运行数据库、快速带入数据库、处理快照和更新管理脚本的功能。当前多模数据库往往存在两个缺陷:一是缺乏统一的访问接口,不能够利用统一查询语言进行查询,比如基于 Azure 云的 Cosmos DB,包含了关系型的 SQL Server,文档型的 Mongo DB,列簇式的 Cassandra 等各种模型存储,但它们都只能用各自对应的 API 进行访问;二是支持统一查询接口的往往查询性能不理想,比如基于文档型的 Arango DB,虽然能够利用统一查询语言 AQL 同时查询文档和遍历图数据,但性能往往不理想,因为利用文档的形式实现图模型需要通过多文档进行连接,当图遍历的深度加大,查询将非常耗时。

未来多模数据库应该是一种原生支持各种数据模型,有着统一访问接口,能自动化管理各模型的数据转化,模式进化且避免数据冗余的新型数据库系统。

## 2.7 云数据库技术

任何信息系统都需要对数据进行计逻辑计算,最终存入数据库中。在云计算中,我们可以把信息系统的数据库简单得分为传统关系数据库、云数据库和 NoSql 数据库。

关于云数据库的特征,主要体现在以下几个方面:

(1) 数据的分布式存储。由于云数据库属于虚拟数据库,并且数据库会根据用户购买云数据库服务等级的不同而制定出不同的模式。通常情况下,云数据库会根据并发连接数、容量的大小,将数据进行拆分并存储到不同的虚拟机当中。并且在拆分的过程中,具有较强的智能算法,实现提升存储效率的目标。

(2) 数据服务的高可用性。传统的数据库具有可用性低的缺陷,而云数据库能够根据不同类型的数据库产品与主机来对存储行为进行选择。通常是采用冗余的机制,就是指在数据处理的过程中拥有多个数据库主机,其主要的目标是能够当出现主机或数据库存在故障时,能够在较短的时间内切换数据服务,因此能够看出云数据库的数据服务具有高可用性特征。

(3) 数据服务的隔离性。当存在多个数据库实例运行在同一台物理主机当中,必然存在数据库间争夺资源的现象,进而严重影响到数据库的服务质量与效率。但是采用云数据库则会将数据库实例进行隔离,由于数据库实例之间不会受影响,因此能够保证其质量与效率。

云与数据库的融合,减少了数据库参数的重复配置,具有快速部署、高扩展性、高可用性、可迁移性、易运维性和资源隔离等特点,具体有两种形态,一种是基于云资源部署的传统数据库,即数据库云服务;另一种是基于容器化、微服



务等理念设计的存算分离架构的云原生数据库。

云原生数据库能够随时随地从多前端访问，提供云服务的计算节点，并且能够灵活及时调动资源进行扩缩容。

## 2.8 HTAP (Hybrid Transaction Analytical Processing)

业务系统的数据处理分为联机事务处理 (OLTP) 与联机分析处理 (OLAP) 两类。针对两类系统的数据管理和系统设计方式都有很大差异：

(1) 对 OLTP 的数据模型采用基本的约束 E-R 图模型，而 OLAP 的数据模型则需要采用特殊的星型模型、数据立方等数据仓库相关的技术。

(2) 对 OLTP 的数据存储通常采用行式组织，而 OLAP 采用列式组织。

(3) OLTP 的业务通常对实时要求比 OLAP 高很多。

(4) 传统的数据库，为了同时支持两类业务。通常采用两个数据源，分别对两套系统进行优化设计。

OLTP 的数据定期会通过 ETL (提取, 转换, 加载) 工具把数据同步导入 OLAP 系统中。这就涉及到数据源滞后的问题。OLAP 的数据滞后，导致分析出来的结果时效性不够，对决策支持类系统的要求不够。

HTAP 是混合 OLTP 和 OLAP 业务同时处理的系统，是一种新兴的应用体系结构，它打破了事务处理和分析之间的界限。广义的 HTAP 数据库是指能够在关系数据模型上进行 OLTP 时具有强一致性保证，并且融合了分布式能力从而同时具有高扩展性。狭义的 HTAP 数据库是指采用行业混存或行列转化同时支持事务处理和联机分析功能的数据库。HTAP 系统主要以下几点优点：数据不需要从操作型数据库导入到决策类系统；操作事务，实时地对分析业务可见；上钻下取等分析操作，时刻操作最新的数据；减少对副本的要求。

目前 HTAP 系统架构大致有如下四种：(1) 行存为主，内存列存为辅；(2) 主机行存、备机内存列存；(3) 多副本行列共存；(4) 是列存为主，行存为辅。

实现 HTAP 的关键技术主要包括行列转换技术、行列共存的查询优化技术、行列共存的事务处理技术等。需要注意的是，HTAP 的价值在于更加简单通用，对于绝大部分中小规模的客户，数据量不会特别大，只需要一套系统即可，但对于超大型互联网企业，HTAP 数据库的分析性能可能不如专用 OLAP 数据库或大数据平台。

## 2.9 数据库与人工智能结合

人工智能与数据库的技术融合可以体现在两个方面：一方面可以通过人工智能技术如机器学习、神经网络等实现数据库的自优化、自监控、自调优、自诊断；另一方面可以实现库内人工智能训练，降低人工智能使用门槛。人工智能与数据库的结合既可以体现在数据库系统自身的智能化，包括数据分布技术智能化、库内进行训练和推理操作、数据库自动诊断、容量预判等，也可以体现在数据库周边工具的智能化，能够在提升管理效率、降低错误引入率、减少安全隐患的同时也大大降低了运营成本。

目前研究重点是将机器学习与数据管理在功能上融合统一，来实现更高的查询和存储效率，自动化处理各种任务。机器学习算法可以分析大量数据记录，标

记异常值和异常模式，帮助企业提高安全性，防范入侵者破坏，还可以在系统运行时自动、连续、无人工干预地执行修补、调优、备份和升级操作，尽可能减少人为错误或恶意行为，确保数据库高效运行、安全无失。

## 2.10 计算存储分离

很多传统数据库是 Shared Nothing 架构，计算资源（CPU）和存储资源（内存、磁盘）都是在一台机器上，优势在于能够降低网络 IO 带来的传输时间成本，最大化提升查询性能。其缺点也非常明显，由于计算资源和存储资源是紧耦合的，所以在能力扩充方面需要同时进行，从而势必产生资源浪费。

在存算分离架构下，存储层和计算层可以根据需求分别进行扩展，一是解决了混合计算存储带来的资源浪费问题，也可以更好地与云平台融合，适应云计算的发展趋势；二是无损容灾将数据库容灾问题转化为更加成熟的分布式存储系统的容灾问题，而类 Paxos 协议在分布式存储中被广泛使用。例如，Snowflake 数据仓库最早提出了独特的存储、计算以及管理服务分离的架构，使得计算层与缓存层并不强耦合。

## 2.11 充分利用新兴硬件

新型硬件的发展为分布式数据库的架构带来了新的变化。数据库系统初始避免数据的搬移，减少网络的 IO 的开销，随着 RDMA 等高速网络的发展，网络 IO 逐渐不再是瓶颈，从存储向计算端搬移数据的效率变快了，数据湖等统一的数据存储层出现了。此外，大内存和高速硬盘渐渐普及，NVM 等非易失内存的发展，很可能颠覆计算机系统的结构，内存计算的潜力会慢慢爆发，分布式数据库也在跟进内存计算的方向。在芯片层，数据库之前以 CPU 为调度核心，随着 GPU、FPGA 和 ASIC 等芯片的发展，以及人工智能、物联网、复杂分析、高性能计算等多样性的任务负载出现，数据库出现了利用异构芯片能力来应对不同任务的趋势。

新兴硬件可以从计算、存储和传输三个层面赋能数据库。计算层面，借助 GPU、FPGA、AI 芯片等，可以实现包括但不限于多核并行优化、事务并发控制、查询加速、存储层计算卸载、数据压缩加速、工作负载迁移等能力；存储层面，随着 NVM 的出现和发展，内存和外存的界限变得模糊，针对传统块存储设计的索引在 NVM 中面临新的性能挑战；传输层面，RDMA 带来网络传输高性能表现和 CPU 卸载能力，为充分榨取利用其性能，可能对数据库系统的架构设计带来颠覆性变化。

非易失性内存（NVM）具有容量大、低延迟、字节寻址、持久化等特性，能够应用于传统数据库存储引擎各个部分，如索引、事物并发控制、日志、垃圾回收等方面；GPU 适用于特定数据库操作加速，如扫描、谓词过滤、大量数据的排序、大表关联、聚集等操作，互联网公司在 FPGA 加速进行了很多探索。

## 3.结语

数据库技术在多年的发展中，取得了巨大的成就。在如今的大数据时代，数据库技术将与新兴技术深度融合，发挥更加重要的作用，成为各个应用领域不可或缺的一部。由于新技术的出现与发展，数据库技术又面临着巨大的挑战。当前数据库呈现多种技术相互融合渗透的趋势，凡是有数据产生的领域就需要数据库

技术的支持，无论是嵌入式移动数据库、微型数据库，亦或是区块链数据库，在新兴领域总可以看到数据库的身影。随着云计算、人工智能、区块链与新型硬件的发展，数据库技术也将不断完善，为社会发展和科技进步贡献更多力量。

### 参考文献：

- [1] 萨师煊，王珊. 数据库系统概论[M]. 高等教育出版社
- [2] 赵希晶. 浅析云数据库的应用现状与展望[J]. 电子技术与软件工程, 2016, (02):185
- [3] 谢红. 基于“云计算”的数据库分析[J]. 科技创新导报, 2011, (14):25-26. .
- [4] 焦通, 申德荣, 聂铁铮, 寇月, 李晓华, 于戈. 区块链数据库:一种可查询且防篡改的数据库[J]. 软件学报, 2019, 30(9): 2671-2685.
- [5] 李俊逸, 王卓, 马鹏玮. 图数据库技术发展趋势研究[J]. 信息通信技术与政策, 2021, 47(05):67-72.
- [6] 唐浩, 姜春宇, 田稼丰. 分布式数据库技术及发展趋势研究[J]. 信息通信技术与政策, 2020, (11):86-90.
- [7] 数据库发展研究报告(2021年). 2021 大数据产业峰会. 2021. 6. 24
- [8] 高瞻, 王飞鹏. 浅谈数据库技术发展趋势[J]. 黑龙江科技信息, 2016, (28):185.